Text-independent Open-set Speaker Identification for Military Missions Using Genetic Rule-based System

Jae C. Oh
Dept. of EECS
Syracuse University
Syracuse, NY 13244
jcoh@ecs.syr.edu

Misty Blowers
AFRL/IFEC
525 Brooks Road, Suite E1
Rome, NY 13441
misty.blowers@rl.af.mil

ABSTRACT

We present a genetic classifier system approach to the text-independent open-set speaker identification problem. Classifier systems are widely used in symbolic problem for dynamically changing open-ended learning. Signal processing problems require processing of real-valued parameters that classifier systems are not designed for. On the other hand, the approaches based on common cepstral encoding with clustering algorithms handle the closed-set speaker identification quite well. This research solves the open-set problem by hybridizing these two approaches.

Categories and Subject Descriptors

I.2.6 [**Learning**]; I.2.7 [**Natural Language Processing**]

General Terms

Design, Algorithms

Keywords

Genetic Classifier Systems, Open-set Text-Independent Speaker Identification

1. INTRODUCTION

We present new *adaptive open-set speaker identification* algorithms using genetic rule-based system [1] for military missions. In a military operation, voices can overlap and have short bursts while the number of speakers can vary over time. Unlike the *closed-set speaker identification problem*, the open-set problem does not assume that the number of speakers is fixed. An open-set speaker identification system must add new speaker profiles as needed dynamically.

The open-set speaker identification is much harder than the closed-set problem because it is hard to decide whether to introduce a profile for a new speaker or to identify a speaker as one of the existing speakers. Commonly, a closed-set speaker identification system utilizes Vector-Quantization combined with a statistical clustering or a neural-network learning. Unfortunately, many of these methods are not suitable for the open-set problem. For example, in clustering-based algorithms, the training phase and the testing phase (i.e., deployment phase) are strictly separated. In order to add a new speaker, a clustering-based algorithm requires reconsideration of all existing speakers with the new speaker. Classifier systems are open learning systems in that training and testing phases are not strictly separated. Therefore, adding new speakers to the system doesn't necessarily require reconsideration of profiles for all existing speakers. We discuss the design and implementation of a new system suitable for the open-set problem. This is a unique methodology to the voice identification problem and it will complement existing methods if not replace them.

2. BACKGROUND

In some situations, it is sufficient to identify a speaker within a closed set of speakers. However, the open set solution is becoming increasingly more important because of its ability to add new speakers.

The human speech production mechanism can be characterized by its components as shown in Figure 1. In this system the air is thrust from the lungs, it passes the vocal cords, passes through the pharynx and then passes out the mouth and nasal cavity [5]. These physical attributes are unique to each person. When a sample of audio data is collected, the computer characterizes it as an audio waveform. The audio waveform can be translated into a physical state of the vocal tract through Linear Predictive Coding [2].

The cepstral features are then generated for each speaker. From this point, existing methods involved subsequently developing a codebook by clustering these snapshots to develop a specific model for each speaker. Once the speaker was characterized in this way, the speaker could be identified from that closed set in the feature. However, if the speaker was not a member of the predefined set, the system forced a decision, and resulted in an erroneous prediction. The new system has the ability to adapt to preexisting speakers natural vocal tract variations (swelling caused by sickness, for example) as well as the ability to update its rule set when a new speaker is presented.

The system learns the rules from the cepstral feature vectors that were mentioned above. LP-Cepstrum (Linear Prediction Derived Cepstrum) is derived from the theory of Ho-

GECCO'05, June 25–29, 2005, Washington, DC, USA.
Copyright 2005 ACM 1-59593-097-3/05/0006 ...$5.00.

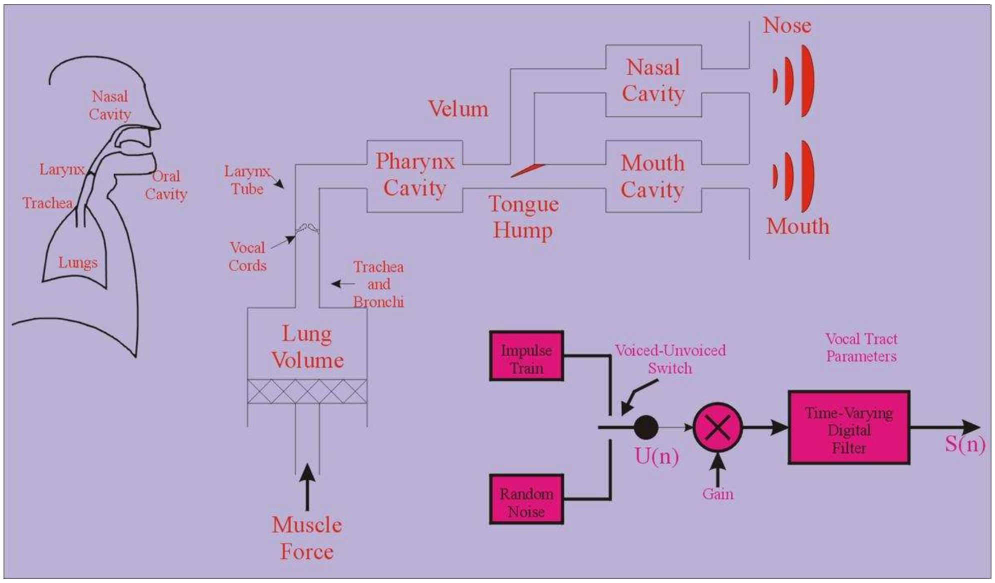

Figure 1: *Human Speech Production Mechanism*

momorphic filtering which aims to map the process of convolution into one of addition and separates the pitch information (fast moving component of speech) from vocal tract characteristic (slowly varying). The system serves to model the slowly varying component of speech [2].

Genetic classifier systems are known to have capability of learning new rules and being an open-ended learning system. However, the traditional classifier systems are not designed for such numerical processing. This is one of the obvious reasons for classifier systems have not been investigated for speech processing problems. Although rare, there are several successful research results on numerical-based problems such as image processing using classifier systems [3, 4]. One of the strengths of the rule-based classifier systems over numerical methods is the ability to utilize temporal information and to use intermediate results as feedback in real-time to refine the decision process. We have developed modifications necessary to the traditional classifier systems for text-independent open-set speaker identification problem.

3. GENETIC RULE-BASED SYSTEM APPLIED TO SPEAKER IDENTIFICATION

Figure 2 shows the genetic rule-based learning system for the SID problem. The system is a typical genetic rule-based learning system except the solid arrows shown. The "if-then" rules in the rule-base have *condition* and *action* part. The condition part of the rule-format for the speaker identification problem looks as follows:

```
<(0.1 2.0), (0.001, 30), (33, 1) (11.1, 11.0002)>
```

and the action part is the name of a speaker in the set of speakers that are identifiable so far.

Each pair represents the interval of a feature from the input. In the example, the first pair `(0.1, 2.0)` says that any input of the first feature between 0.1 and 2.0 will match with the first condition. The second means that any input between 0.001 and 30 will match with the second condition. The third pair is don't-care condition because the first number is larger than the second. This means that this particular rule doesn't consider the third feature from the input making decisions. The action part, `John` represents that any input matching with the rule is identified as John.

The rule encoding is done in a way that the learning component will be able to emphasize relevant features of each speaker but to ignore features that are not important.

In the figure, *Step 1* shows the input to the system that is the feature vector of the speaker which is converted to detector messages. *In Step 2*, the detector messages are compared against the condition part of the rules in the rule-base. *In Step 3*, the rules that match their condition part with one or more detector messages will post their action part to the Effector Messages list with their *bidding* amounts. A bidding amount of a posting rule represents how confident the posting rule is about its action part. *In Step 4,* a conflict resolution is performed if there are more than one posts and if the suggested actions from the postings are different. Finally, a response to the outside world is made in *Step 5* as an attempt to identify the speaker. The environment, in turn, gives positive or negative feedback depending on the current action of the system (Step 8). In Step 7, learning is performed.

The effector-condition match loop (Step 9) matches the current effector messages to the condition part of the rules without going through the detector message list. This mechanism is useful during the operations of triggered-learning mechanism. Intuitively, this inner-loop attempts to refine rules without accepting new training data instances. The inner-loop can occur more than once, as many times as necessary before exiting to the outer-loop.

There are other necessary modifications of the traditional classifier system to make it work with the SID problem such as specialized genetic operators, fitness functions, and temporal rule-matching method. Due to the space limit, we defer these concepts to another paper.

If the identification of the current speaker fails, the system will regard the speaker as a new speaker and add necessary

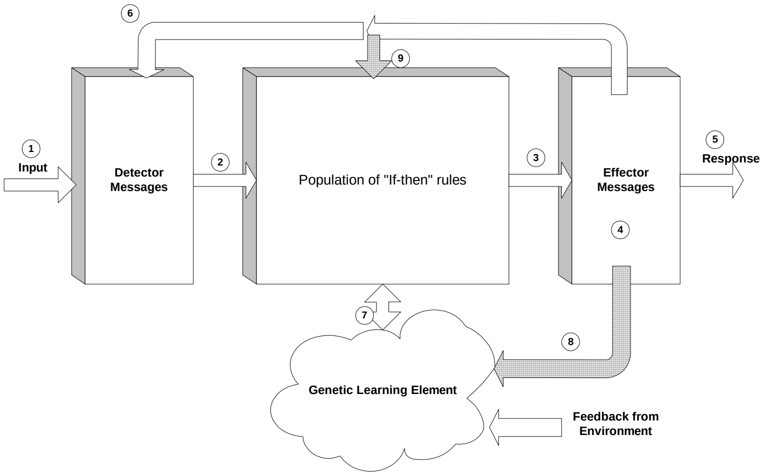

Figure 2: *Block diagram of Genetic Rule-based Classifier System Proposed for Speaker Identification*

classifiers for the speaker. One of the methods we are currently considering is *rule-covering* [6]. Rule-covering creates matching classifiers for the unrecognized input conditions– i.e., new feature vector.

4. EXPERIMENTAL RESULTS

The voice feature vector consists of 14 real numbers generated by the sampling method mentioned above. To learn and identify a speaker, the speaker's voice is sampled and feature vectors are given one by one to the system. The feature vectors are given until the speaker is identified. With a two-speaker problem, the system was able to learn and identify the speakers within 20 iterations on average over 100 experiments. We are currently conducting experiments with more number of speakers. Preliminary results show that the system learns over ten speakers without any significant difficulties.

5. FUTURE WORK AND SUMMARY

We discussed a new approach to the text-independent open-set speaker identification problem using genetic classifier system for adaptation, robustness, and open-ended learning. Unlike many existing methods based on statistical clustering, the new system can dynamically learn new speakers even after deployment. Currently, we have tested the system with identification tasks with relatively small number of speakers. We are improving the system so that it will be able to learn more than 100 speakers.

Acknowledgments

This research was conducted through a grant provided to Syracuse University and through ongoing research in the speaker identification problem at the Air Force Research Laboratory's Information Directorate (AFRL/IFEC), Rome, NY. Special thanks to John Mauew and Matt Wolf for implementing and experimenting with the SID classifier system.

6. REFERENCES

[1] J. H. Holland and J. S. Reitman. Cognitive systems based on adaptive algorithms. In Waterman and Hayes-Roth, editors, *Pattern-Directed Inference Systems*, pages 313–329. Academic Press, 1978.

[2] R. Mammone. *Computational Methods of Signal Recovery and Recognition*. John Wiley & Sons, Inc., New York, 1992.

[3] A. D. McAulay and Jae C. Oh. Improved learning in genetic rule-based classifier systems. In *IEEE International Conference on Systems, Man, and Cybernetics*, pages 1393–1398. IEEE, 1991.

[4] W. Niblack, D. Petkovic, and D. Damian. Experiments and evaluations of rule based methods in image analysis. *CVPR*, 88:123–128.

[5] L. Rabiner and B. Juang. *Fundamentals of Speech Recognition*. Prentice-Hall, New York, 1993.

[6] Stewart W. Wilson. Get real! xcs with continuous-valued inputs. In *Learning Classifier Systems, From Foundations to Applications*, pages 209–222, London, UK, 2000. Springer-Verlag.