Signal Processing 126 (2016) 125–133

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Nonlinear signal processing for vocal folds damage detection based on heterogeneous sensor network

Zhen Zhong [a], Baoju Zhang [b], Tariq S. Durrani [c], Shuifang Xiao [a]

[a] Department of Otolaryngology Head and Neck Surgery, Peking University First Hospital, Beijing 100034, China
[b] College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin 300387, China
[c] Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, Scotland, UK

ARTICLE INFO

Article history:
Received 1 August 2015
Received in revised form
20 August 2015
Accepted 24 August 2015
Available online 14 September 2015

Keywords:
Heterogeneous sensor network
Interval type-2 fuzzy logic systems
Bayesian classifier
Vocal folds
Short-time-Fourier-transform
Singular-value decomposition

ABSTRACT

Heterogeneous sensor network-based medical decision making could facilitate the patient diagnosis process. In this paper, we present an intelligent approach for vocal folds damage detection based on patient's vowel voices using heterogeneous sensor network. Based on human voice samples and Hidden Markov Model, we show that transformed voice samples (linearly combined samples) follow Gaussian distribution, further we demonstrate that a type-2 fuzzy membership function (MF), i.e., a Gaussian MF with uncertain mean, is most appropriate to model the transformed voices samples, which motivates us to use a nonlinear signal processing technique, interval type-2 fuzzy logic systems, to handle this problem. We also apply Short-Time-Fourier-Transform (STFT) and Singular-Value-Decomposition (SVD) to the vowel voice samples, and observe that the power decay rate could be used as an identifier in vocal folds damage detection. Two fuzzy classifiers, a Bayesian classifier and a linear classifier, are designed for vocal folds damage detection based on human vowel voices /a:/ and /i:/ only, and the fuzzy classifiers are compared against the Bayesian classifier and linear classifier. Simulation results show that an interval type-2 fuzzy classifier performs the best of the four classifiers.

1. Introduction

The proliferation of Heterogeneous Sensor Networks (HSN) has created a large amount of multi-sensor signals across multi-modality (e.g., optical, EO/IR, acoustic/seismic, RF, electromagnetic, mechanical, thermal, and electrical). Humans display a remarkable capability to perform multi-modal signal processing despite noisy sensory signals and conflicting inputs. Humans are adept at network visualization, and at understanding subtle implications among the network connections. To date, however, human's innate ability to process and integrate signals from disparate, network-based sources has not translated well to signal processing in HSN. One of the great mysteries of the brain is cognitive control. How can the interactions between millions of neurons result in behavior that is coordinated and appears willful and voluntary? There is consensus that it depends on the prefrontal cortex (PFC) [1–3]. Many PFC areas receive converging inputs from at least two sensory modalities [4,5]. For example, the dorsolateral (DL) and ventrolateral PFC both receive projections from visual, auditory, and somatosensory cortex. Furthermore, the PFC is connected with other cortical regions that are themselves sites of multimodal convergence. The Stroop task, naming the color of a conflict stimulus (e.g., the word GREEN displayed in red) [6,8], and Wisconsin card sort task (WCST) [9] are variously described as tapping the cognitive functions of either selective attention, behavioral inhibition, top-down control, working memory, or rule-based or goal-directed

E-mail addresses: zhong-zhen@sina.com (Z. Zhong), wdxyzbj@163.com (B. Zhang), durrani@strath.ac.uk (T.S. Durrani), xiao-ent@163.com (S. Xiao).

http://dx.doi.org/10.1016/j.sigpro.2015.08.019
0165-1684/

behavior [1]. In this paper, we are interested in rule-based approach for decision making based on HSN. In [7], opportunistic sensing for HSN was studied for two different cases, with independent modalities and correlated modalities. It showed that HSN with independent modality sensors, we can make analysis and information fusion for each single modality sensor network. On decision making based on sensor networks, many interesting works have been reported, for example, target recognition [10–12], sense-through-wall human detection [13], threat assessment [15], resource allocation [16], and localization [17]. In this paper, we are interested in an important medical decision making problem, vocal folds damage detection, based on heterogeneous sensor networks with optical and acoustic modalities.

In decision theory, *ambiguity* about probabilities should not affect choices. However, recent experiments [14] showed that many people are more willing to bet on risky outcomes (e.g., gambling on a roulette wheel) than on an ambiguous one (e.g., chance of a terrorist attack based on meager or conflicting evidence), holding the judged probability of outcomes constant. So, the confidence in judged probability can vary widely for "risky" and "ambiguous". Using functional brain imaging, Hsu et al. [14] showed that the *level of ambiguity* in choices correlates positively with activation in the amygdala and orbitofrontal cortex, and negatively with a striatal system. This suggests that *degree of uncertainty* should be considered in decision making, contrary to traditional decision theory. So we introduce degree of uncertainty into vocal folds damage detection. The degree of uncertainty will be represented using type-2 fuzzy logic.

The presence of any form of damage to the delicate vocal fold cover tissue will impair phonation, be it a polyp, nodule, edema or scarring. Polyps of the vocal folds are a separate entity occurring nowhere else in the larynx or in the human body [18]. The pathological diagnosis of the vocal folds is a field which demands more investigation. Traditionally, the methods of diagnosis are indirect laryngoscope, video-laryngoscope, and stroboscope light [19]. However, most of these methods need special instrument, and mainly depend on the experience of the pathologists. In [20], measurement of vocal fold intraglottal pressure and impact stress were studied, and the experimental results match well with analytical predictions and support a current theory of mechanical trauma leading to vocal nodules. In [21], a nonlinear model was proposed to study chaotic vibrations of vocal folds with a unilateral vocal polyp.

Many engineering approaches for voice pathology detection have been reported. In [22], direct speech feature estimation using an iterative Expectation–Maximization (EM) algorithm was applied to vocal fold pathology detection. In [19], a measure of pitch perturbation was shown to be most useful in pathological discrimination. A robust, rapid and accurate system for automatic detection of normal and pathological speech was studied in [23], which employed non-invasive, non-expensive and fully automated measures of vocal tract characteristics and excitation information. In [24], jitter estimation over short time intervals (short-term jitter) was used for voice pathology detection in the case of running or continuous speech, where short-term jitter estimations were provided by the spectral jitter estimator, which was based on a mathematical description of the jitter phenomenon. In [25], a system for remotely detecting vocal fold pathologies using telephone-quality speech was presented, which used a linear classifier, processing measurements of pitch perturbation, amplitude perturbation and harmonic-to-noise ratio derived from digitized speech recordings. In [26], a neural network-based detector processing Mel frequency cepstral coefficients and their derivatives for normal/abnormal discrimination was presented. In [42], some preliminary results on vocal folds damage detection were presented based on patient voices. In [43], throat polyp detection based on compressed big data of voice with support vector machine algorithm was presented, and in [44] an intelligent throat polyp detection approach with separable compressive sensing was proposed.

In this paper, we are interested in detecting vocal folds damage using acoustic sensors based on the patient vowel voices only. Traditional pattern recognition techniques such as Bayesian classifier, known as the optimal classifier, could be used if the voice samples follow certain distribution, and this belongs to model-based statistical processing. As noted in [27], a shortcoming to model-based statistical signal processing is "…the assumed probability model, for which model-based statistical signal processing results will be good if the data agrees with the model, but may not be so good if the data does not." In human's voices, the voice amplitude is highly bursty, and we believe that no statistical model can really demonstrate the uncertain nature of the voice. Fuzzy logic systems (FLS) are model free. Their membership functions are not based on statistical distributions. In this paper, we, therefore, apply fuzzy techniques to vocal folds damage patient diagnosis.

In Section 2, we model voice samples using interval type-2 Gaussian membership function. In Section 3, Short-Time Fourier Transform and Singular-Value Decomposition are used to identify patient voice. Two fuzzy classifiers are presented in Section 4. In Section 5, a Bayesian classifier is proposed. Performances of four classifiers (including an existing linear classifier approach) are evaluated in Section 6. Conclusions are presented in Section 7.

2. Modeling voice samples using hidden Markov model and Gaussian primary MF with uncertain mean

In [39], autoregressive Hidden Markov Model (HMM) was used to represent voice samples x_i, which means we could have

$$x_k = -\sum_{i=1}^{p} b_i x_{k-i} + n_k \tag{1}$$

where n_k is Gaussian noise, and b_i $(i = 1, 2, \ldots, p)$ are the autoregression coefficients where p is autoregressive order. So

$$x_k - \sum_{i=1}^{p} c_i x_{k-i} = n_k \tag{2}$$