www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 4 April 2017, Page No. 20903-20905
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i4.17

Analysis of Speech Disabled People with Dysarthria using TORGO Database

Usha.M
Assistant Professor & Head,
Department of Computer Applications,
K G College of Arts & Science

Abstract— This paper represents a method for assessment of speech skill of dysarthria disordered patients repeatedly, which is a motor speech disorder comes from neurological injury of the motor element of the motor-speech system. This structure comprises two major features: feature representation and prediction. In the feature representation, the speech sound is changed into a phone series with an regular speech recognition technique and joint with a canonical phone sequence from an intonation dictionary using a biased finite state transducer to obtain the pronunciation mappings such as match, substitution, and deletion. Next, in the prediction process, a structured sparse linear model exists with phonological information that concurrently addresses phonologically structured sparse feature selection and transparency prediction is identified. The TORGO database of dysarthric pronunciation consists of aligned acoustical and measured 3D articulatory features from speakers with either cerebral palsy (CP) or amyotrophic lateral sclerosis (ALS), which are two of the most common causes of speech disability. The major aim of this paper is to supply a flexible and suitable environment for the dysathria speech disordered peoples by recognizing their speech to make others understand them

.Keywords—dysarthria, feature extraction, prediction

I. INTRODUCTION

Speech production is one of the main "impressive motor skills". Control of speech movements follows a way of development up to age 12; humans acquire adult-like speech motor control by adolescence Childhood motor speech problems most likely caused by neurological difficulties, and adults can experience also after injuries or illnesses.

Two major categories are apraxia and dystarthria. Speech production shortage that results from impairment of the neuromuscular and/or motor control system. Patients disorder may co-occur with other language impairments. Dysarthria Patients have troubles in controlling the motor subsystems, including respiration, phonation, articulation, and prosody [1]. The speech lucidity of that people is reduced in proportion to the harshness of dysarthria, and patients with dysarthria have complexity in communicating with others [2].

In the experimental scenario, one of the significant goals of treatment is to rise the improvement in articulation because increased articulation, clarity of phonemes is the best essential factors affecting speech perception [3].

The amount of speech therapists is very little compared to that of patients with dysarthria. Also, an objective automatic assessment with a computational algorithm, which may assist or replace an expert investigation, is cost-effective, bias-free, flexible, and repeatable. Hence, an automatic process that is highly correlated with the perceptual scores given by an expert will be useful in assisting clinicians in the diagnosis and treatment of speech disorders.

An important issue in promoting a reliable, objective assessment process is how to extract and select proper speech characteristics that capture distinctive features depending on the intelligibility of dysarthric speech [4].

II. FEATURE REPRESENTATION AND PREDICTION

Feature Representation

In this process, the speech utterance is converted into a phone sequence using an automatic speech recognition technique and is aligned with a canonical phone sequence from a pronunciation dictionary using a weighted finite state transducer to recognize the pronunciation mappings such as match, substitution, and deletion.

Weighted Finite State Transducers (WFST)

A finite-state transducer is a finite automaton whose state transitions are represented with both input and output symbols. Therefore, a path through the transducer encodes a mapping from an input symbol sequence, or string, to an output string.

A weighted transducer joins weights on transitions in addition to the input and output symbols. Weights may encode probabilities, durations, penalties, or any other quantity that accumulates along paths to compute the overall weight of mapping an input string to an output string. Weighted transducers are thus a natural choice to represent the probabilistic finite-state models prevalent in speech processing. Finite State Transducers can be weighted, where each transition is labelled with a weight in addition to the input and output levels. A Weighted Finite State Transducer (WFST) over a set K of weights can be defined similarly to an unweighted one as an 8-tuple $T = (Q, \Sigma, \Gamma, I, F, \mathrm{E}, \lambda, \rho)$, where:

Q, Σ, Γ, I, F is defined as above;

Maps initial states to weights;

Maps final states to weights.

Finite-State models (FSM) and, specifically, Weighted Finite-State transducers (WFST) have demonstrated very fruitful in many fields of written and spoken language processing. This incorporates in particular machine translation, large vocabulary continuous speech recognition and speech synthesis. An

intriguing component of FSMs is that they can be automatically built or "learned" from training data using corpus based techniques.

Compared to the more traditional knowledge based approaches, these techniques are attractive for their potential of much lower development cost. Another interesting property of FSMs is their feasibility for implementing or approximating knowledge-based techniques.

Different knowledge sources can hence be represented via FSMs, thus allowing the integration of apriori knowledge with inductive techniques in a natural and formally elegant way. This makes the FSM framework an adequate one for language processing. The main goal of this project is the application of this framework to speech recognition and synthesis, which will constitute the themes of the two major tasks.

The group has already acquired some experience in modeling the various components of recognition systems using WFSTs, having developed specialized algorithms for transducer composition for on-the-fly lexicon and language model integration.

Preliminary experiments with the explicit integration of phonological rules have also produced encouraging results, but much remains to be done in order to be able to achieve higher recognition rates, especially in what concerns spontaneous speech, not only in terms of pronunciation modeling, but also of language modelling.

The group's experience with the application of WFSTs to various modules of text-to-speech (TTS) synthesis is substantially more later. We are currently examining the capabilities of WFST application, namely in terms of grapheme-to-phone conversion and variable duration segment selection in synthesis by concatenation. We plan to proceed with this preliminary work and stretch out it to other modules.

A third task will deal with more exploratory themes on which the group currently has no experience, having as a goal the integration of different knowledge sources. Potential themes are the use of transducers for topic indexation, or for integrating translation models to help speech recognition of documents which have, originally been written in one language and are being dictated by a human translator in another language.

Prediction

Sparse linear model used to find the relevancies between the sound utterance and the phonological knowledge required more computation overhead due to thepresence of the large volume of data.

Sparse Linear Model

Sparse Linear Models Consider the general problem of regression, including variables $y \in R\ n$, $X \in R\ n \times p$, $\beta \in R\ p$ (sparse vector), $\in R\ n$.

We write a linear model with p features and n samples as: $y = X\beta + e$. Recall that, in this linear regression model,

y are the response terms; a noisy version of Xβ

X is the observed predictors or covariates

β coefficients (unobserved), weigh each of the p features in X depending how well feature j can be used to predict Y.

III. TORGO DATABASE

The TORGO database of dysarthric articulation comprises of related acoustics and measured 3D articulatory characteristics from speakers with either cerebral palsy (CP) or amyotrophic lateral sclerosis (ALS), which are two of the most imperative purposes for speech disability, and matched controls.

Both CP and ALS effect in dysarthria, which is caused by disruptions in the Neuro-motor interface. These disruptions distort motor commands to the vocal articulators, resulting in a typical and relatively unintelligible speech in most cases. This unintelligibility can radically diminish the use of traditional automatic speech recognition (ASR) software. The incapability of modern ASR to effectively understand dysarthric speech is a major problem, since the more general physical disabilities often connected with the condition can make other forms of computer input, such as keyboards or touch screens, especially difficult

The TORGO database was initially primarily a resource for developing advanced models in ASR that are more appropriate to the needs of people with dysarthria, although it is also applicable to non-dysarthric speech. A main reason for collecting detailed physiological information is to be able to explicitly learn 'hidden' articulatory parameters automatically via statistical pattern recognition. For instance, recent research has shown that modelling conditional relationships between voice and acoustics in Bayesian networks can reduce error by about 28% relative to acoustic-only models for regular speakers.

Each speaker is assigned a code and contributed their own directory. Female speakers have a code that starts with 'F' and male speakers have a code that starts with 'M'. If the speaker is a member of the control group (i.e., they do not have a type of dysarthria), then the letter 'C' takes after the gender code. The last two digits merely indicate the order in which that subject was enlisted. For example, speaker 'FC02' is the second female speaker without dysarthria selected.

Each speaker's directory contains 'Session' directories, which typify information recorded in the respective visit, and occasionally a 'Notes' directory which can include Frenchay assessments, notes about the sessions (e.g., sensor errors), and other significant notes.

Each 'Session' directory can contain the accompanying content:

alignment.txt	This is a text file containing the sample offsets between audio files recorded simultaneously by the array microphone and the head-worn microphone. The first line is a space-separated pair of directories indicating that indicated offsets refer to files in the second directory relative to those in the first. All subsequent lines in alignment.txt indicate the common filename and the sample offset, separated by a space.
amps/	These directories contain raw *. amp and *. in files produced by the AG500 articulograph.
phn_*/	These directories contain phonemic transcriptions of audio data. Each file is plain text with a *.PHN file extensions and a filename referring to the utterance number. These records were produced using the free Wavesurfer tool according to the TIMIT phone set, with phonemes marked *cl referring to closures before plosives. Files in 'phn_arrayMic' are aligned temporally with acoustics recorded by the array microphone and files in 'phn_headMic' are adjusted temporally

with acoustics recorded by the head-worn microphone.

pos/ These directories contains the head-corrected positions, velocities, and orientations of sensor coils for each utterance, as created by the AG500 articulograph. These documents can be read by the 'loaddata. m' function in the included 'tapadm' toolbox (see below) and contain the essential articulatory information of interest. Except where noted, the channels in these data allude to the following positions in the vocal tract:
Tongue back (TB)

prompts/ These directories contain orthographic transcriptions. Each filename refers to the utterance number. Prompts checked 'xxx' demonstrate spurious noise or otherwise generally unusable content. Prompts having a *. JPG file refers to images in the Webber Photo Cards: Story Starters collection.

rawpos/ These directories are equal to the post/ directories aside from that their articulographic content is not head-normalized to a steady upright position.

wav_*/ These directories contain the acoustics. Each file is a RIFF (little-endian) WAVE audio file (Microsoft PCM, 16 bit, mono 16000 Hz). Filenames allude to the utterance number. Files in 'wav_arrayMic' are recorded by the array microphone and files in 'wav_headMic' are recorded by the head-worn microphone.

IV. CONCLUSION AND FUTURE ENHANCEMENTS

The main goal of this system is to provide a flexible and convenient environment for the dysathria speech disordered TORGO Database is implemented for the assessment of Dysartria speech disabled people. peoples by recognizing their speech to make others understand them. In this paper, assessment of dysathria speech, so that others can understand it efficiently.

In future, the accuracy of speech recognition can be improved by introducing the SVM Based Learning Methodology which can classify the dysathria speech based on which more relevant matching can be done.

References

[1] J. R. Duffy, Motor Speech Disorders: Substrates, Differential Diagnosis, and Management. St Louis, MO, USA: Elsvier Mosby, 2005.

[2] H. Kim, K. Martin, M. Hasegawa-Johnson, and A. Perlman, "Frequency of consonant articulation errors in dysarthric speech," Clinical Linguist. Phonet., vol. 24, no. 10, pp. 759–770, Oct. 2010.

[3] R. Palmer and P. Enderby, "Methods of speech therapy treatment for stable dysarthria: A review," Adv. Speech Lang. Pathol., vol. 9, no. 2, pp. 140–153, 2007.

[4] L. D. Shriberg and J. Kwiatrkowski, "Phonological disorders III: A procedure for assessing severity of involvement," J. Speech Hear. Disorders, vol. 47, no. 3, pp. 256–270, 1982.

[5] G. Davis, "Noise Reduction in Speech Applications", Electrical Engineering & Applied Signal Processing Series, CRC Press. Publication, 2002.

[6] Chunshien Li, "Soft computing approach to adaptive noise filtering", Cybernetics and Intelligent Systems, 2004 IEEE Conference.

[7] Boll, S., "Suppression of acoustic noise in speech using two microphone adaptive noise cancellation", Acoustics, Speech and Signal rocessing, IEEE Transactions on (Volume: 28, Issue: 6).

[8] Rudzicz, F., Namasivayam, A.K., Wolff, T. "The TORGO database of acoustic and articulatory speech from speakers with dysarthria", Language Resources and Evaluation, pp. 1-19, 2011.