On the Predictive Power of Acoustic Features in the Automatic Assessment of Dysarthria

Nasim Mahdinazhad Sardhaei, Stefan Werner

Department of Language Studies, University of Eastern Finland, Joensuu, Finland

Introduction

Speech evaluation in Motor disorders can have a prominent role in obtaining valuable pieces of information on the disturbances to the neuro muscular mechanisms affecting the speech production by providing objective measures. Many researchers have verified the efficiency of speech assessment as a promising tool providing prognosis markers to discriminate healthy and pathological patterns (Meilán et al., 2014; Ash & Grossman, 2015; Poole et al., 2017). From another perspective, advances of technology in computer sciences have added a new dimension to the processing of impaired speech leading to capture broad cues of the speech disorders. More specifically, the development of various speech processing technologies has paved the way to examine diverse methods of automatic assessment of the symptoms as well as the severity of speech impairments. Speech classification systems as an embodiment of speech technologies involve the automatic classification of input audio signals and prioritization of the most relevant signals. The feature extraction, as an important stage of developing a classification system, primarily focuses on the extraction of wide range of acoustic features which can grasp the existing variabilities in pathological speech patterns or sources. At this stage, the acoustic based methods adopting machine-learning approaches automatically extract a large number of measurement features and subsequently the models are learned from the large sets of data resulting in an automatic speech classification system. Despite having several advantages, one of the most crucial drawbacks of such methodology using to multi-parameter models is it prevents the possibility to gain explicit knowledge on the most relevant acoustic parameters. With high dimensionality of feature vectors, the recognition of the relevant and irrelevant features becomes more challenging and consequently some irrelevant features containing less useful information may bring errors into the classification system. As not all features contribute positively to the performance of a classifier, it becomes of great importance to remove the source of error. A direct solution to the above-mentioned issues caused by high feature dimensionality can be reducing the dimensionality of feature vectors.

Objectives

The present study aims at scrutinizing the strength of acoustic analysis for the objective assessment of dysarthria.

More specifically, the focus of the current study lies on the reduction of the feature space of multi-parameter models and search for an optimal subset of acoustic parameters from very large number of acoustic features derived automatically from dysarthric speech.

It is hypothesized that the reduced parameter set will maintain efficient predictive power to automatically detect the severity of dysarthria by means of *linear statistics* and predict the response variable which is the Frenchay (FDA) severity score.

Methodology

The corpus used in this research was the Nemours Database of Dysarthric Speech consisting of the speech data of 11 American male speakers with varying degrees of dysarthria each producing 74 short nonsense sentences with restricted vocabulary containing 4 to 6 words. The score from the Frenchay Dysarthria Assessment (Enderby, 1983) for each participant is included in the database. The feature set for the 2012 Interspeech Speaker Trait Challenge (Schuller et al. 2012) from openSMILE toolkit (Eyben et al. 2013) was used for the extraction of acoustic features containing a total of 6125 features derived from 64 energy-, spectrum- and voicing-related "low-level descriptors" (LLD).

A preselection of the features was carried out at the very first stage of the process. The acoustic features were categorized into separate groups based on sharing statistical functionals applied to LLDs in which the features with the functionals arising from the same or similar sources were included into the same group. Then, the Akaike information criterion (IAC) was calculated for each of the members within the group and one or two features with the lowest Akaike information criterion (AIC) were retained from every group.

The stepwise selection was performed on the group of retrained features to find the best combination of the features. The stepwise algorithm produced a predictive model of the Frenchay assessment scores with the smallest AIC= 2811.19 and the combination of the 238 survived variables. Finally, multiple regression analysis was used to validate the effectiveness of the model as a good fit and the results of the multiple regression analysis revealed the linear model is significant $F(238,501)=183.5$, $p<0.001$) with the $R2=0.9833$ showing that 98% of the variance in the data can be explained by the constructed linear model.

Methodology

4 energy related LLD
Sum of auditory spectrum (loudness) Sum of RASTA-style filtered auditory spectrum RMS Energy Zero-Crossing Rate
54 spectral LLD
RASTA-style auditory spectrum, bands 1-26 (0–8 kHz) MFCC 1–14 Spectral energy 250–650 Hz, 1 k–4 kHz Spectral Roll Off Point 0.25, 0.50, 0.75, 0.90 Spectral Flux, Entropy, Variance, Skewness, Kurtosis, Slope, Psychoacoustic Sharpness, Harmonicity
6 voicing related LLD
F0 by SHS + Viterbi smoothing, Probability of voicing logarithmic HNR, Jitter (local, delta), Shimmer (local)

64provided Low-level descriptors 'LLD' (Schuller et al. 2012).

Functionals applied to LLD / Δ LLD
quartiles 1–3, 3 inter-quartile ranges 1 % percentile (≈ min), 99 % percentile (≈ max) position of min / max percentile range 1 %–99 % arithmetic mean, root quadratic mean contour centroid, flatness standard deviation, skewness, kurtosis rel. duration LLD is above / below 25 / 50 / 75 / 90% range rel. duration LLD is rising / falling rel. duration LLD has positive / negative curvature gain of linear prediction (LP), LP Coefficients 1–5 mean, max, min, std. dev. of segment length
Functionals applied to LLD only
mean of peak distances standard deviation of peak distances mean value of peaks mean value of peaks – arithmetic mean mean / std.dev. of rising / falling slopes mean / std.dev. of inter maxima distances amplitude mean of maxima / minima amplitude range of maxima linear regression slope, offset, quadratic error quadratic regression a, b, offset, quadratic error percentage of non-zero frames

Applied functionals (Schuller et al. 2012).

Results

The acoustic survived features selected by the stepwise regression in the final model were classified according to their LLD groups which are presented in the table A. Furthermore, the functions applied to the features were also categorized on the basis of the belonging group and their occurrence size (table. B). The results revealed that the most occurring feature was *audSpec_Rfilt_sma.* (RASTA-style filtered auditory spectrum) from the energy related LLD. Another main feature appearing 49 times in the model was *pcm_fftMag_mfcc_sma.* (MFCC 1-14) which was a member of spectral group. The acoustic feature *audspec_lengthL1norm_sma_* (Sum of auditory spectrum) from the energy related group and the feature *pcm_fftMag_spectralRollOff* 0.***Number*** (Spectral Roll Off Point 0.25, 0.50, 0.75, 0.90) from the group of spectral LLD each had appeared 10 times in the final model. The occurrence size of other features in the model was less than 10 times.

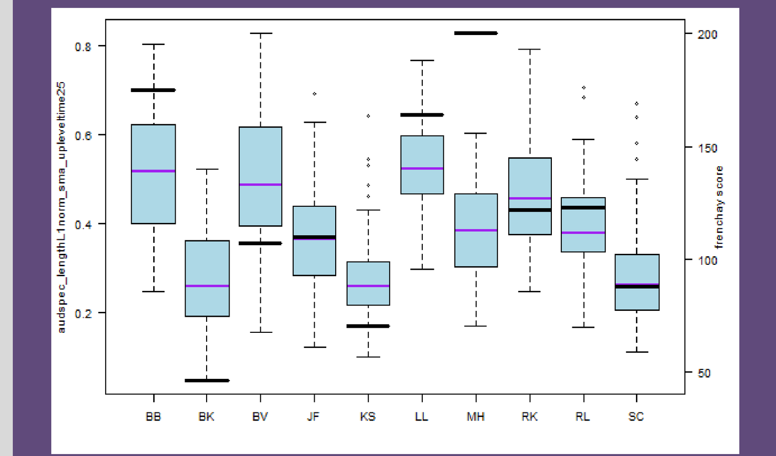
Two example boxplots of acoustic features involved in the generated significant linear model can be observed in Figures 1 and 2. where the features are close to the top of the candidate ranking. As clear in all the figures, the scores of the Frenchay assessment (right x-axis) are overlaid across many subject-specific boxplot distributions of the feature values (left y-axis). Another finding noticeable in all figures is the relation between the relative variations in the severity of the disorder and the feature value distributions which is in line with the work of Werner (2018) recognizing that many relative severity differences are represented in the distributions of the acoustic feature values.

Energy related LLD	Size
Sum of auditory spectrum (loudness)	10
Sum of RASTA-style filtered auditory spectrum	92
RMS Energy	5
Zero-Crossing Rate	7
Spectral LLD	
RASTA-style auditory spectrum, bands 1-26	5
MFCC 1–14	49
Spectral energy 250–650 Hz	1
Spectral energy 1 k–4 kHz	5
Spectral Roll Off Point 0.25, 0.50, 0.75, 0.90	10
Spectral Flux, Entropy	11(5+6)
Spectral Variance	5
Spectral Skewness and Kurtosis	8(4+4)
Spectral Slope	5
Spectral Harmonicity	4
Spectral Sharpness (Psychoacoustic)	6
Voicing related LLD	
F0 by SHS + Viterbi smoothing	4
Jitter (local, delta)	3(2+1)
Shimmer (local)	5
Probability of voicing logarithmic HNR	2

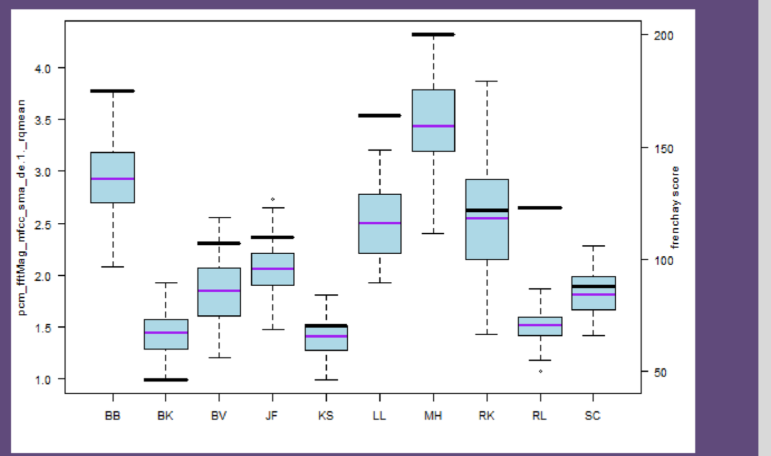
A. Classification of the main features involved in the optimal mode.

Group	Feature	Size
1	Range, iqr1.2, iqr1.3, quartile1, quartile2, quartile3, percentile1.0, percentile99.0, pctlrange0.1, stddev, kurtosis	63
2	minSegLen, maxSegLen, meanSegLen	30
3	upleveltime25, upleveltime50, upleveltime90, risetime, falltime	28
4	lpgain, lpc3	34
5	amean, flatness, rqmean, posamean	29
6	meanPeakDist, peakDistStddev, peakRangeAbs, peakMeanAbs, peakMeanRel, minRangeRel, peakMeanMeanDist,	20
7	meanRisingSlope, stddevFallingSlope	15
8	linregc2, linregerrQ	14
9	qregerrQ	13

B. Classification of the main feature functions involved in the optimal mode.



1. Subject-specific boxplots of the feature variable 'audspec_lengthL1norm_sma_upleveltime25'. Black lines represent the subjects' total Frenchay scores across all ranges of feature values.



2. Subject-specific boxplots of feature variable 'pcm_fftMag_mfcc_sma_de.1._rqmean"

Conclusion

The overall findings revealed a trend supporting the hypothesis that acoustic measures are potentially usable for predicting the dysarthria severity. The stepwise regression analyses of a high dimensional feature set containing the features ranked on the basis of smallest AICs yielded the results which verify the efficiency of the utilized statistical procedure as an alternative solution for the problem of high dimensionality in feature set. Subjecting parameters of the reduced features sets into linear multiple regression analyses confirmed the significance of the model for predicting the Frenchay severity scores.Notwithstanding, construction of an automatic evaluation of pathological speech is a very complex task and much more research is needed to get insights into objective acoustic correlates of dysarthria. Also, in spite of the fact that reduced set of parameters were capable of predicting the severity scores of the dysarthria, relying exclusively on individual variables will not generate a highly reliable automatic assessment. Thus, for having a set of acoustic features with full predictive power, it is also of a great importance to seek for feasible procedures to inspect the interaction between acoustic parameters.