STATISTICAL ANALYSIS OF THE SIGNAL AND PROSODIC SIGN OF COGNITIVE IMPAIRMENT IN ELDERLY-SPEECH
A Preliminary Study

Shohei Kato, Yuta Suzuki
Dept. of Computer Science and Engineering, Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan

Akiko Kobayashi, Toshiaki Kojima
Ifcom. Co., Ltd.
Time24 Bldg. 4F, 2-45, Aomi, Koto-Ku, Tokyo 135-8073, Japan

Hidenori Itoh
Nagoya Institute of Technology, Nagoya, Japan

Akira Homma
Tokyo Dementia Care Research and Training Center
1-12-1, Takaido Nishi, Suginami-ku, Tokyo 168-0071, Japan

Keywords: Early detection of dementia, Speech prosody, Acoustic analysis, Feature selection, Speech prosody-based cognitive impairment rating.

Abstract: This paper presents a novel approach for early detection of cognitive impairment in the elderly. Our approach incorporates the use of speech sound analysis and multivariate statistical techniques. In this paper, we focus on the prosodic features of speech. Fifty six Japanese subjects (22 males and 34 females between the ages of 64 and 90 years) participated in this study. Blind to clinical groups, we collected speech sounds from segments of dialogue during an HDS-R examination. The segments corresponds to speech sounds from answers to questions about time orientation and number backward counting. Ninety eight prosodic features were extracted from each of the speech sounds. These prosodic features consisted of spectral and pitch features (13), formant features (61), intensity features (22), and speech rate and response time (2). These features were refined by principal component analysis and/or feature selection. In addition, we calculated speech prosody-based cognitive impairment rating (SPCIR) by multiple linear regression analysis. The results indicate that a moderately significant correlation exists between the HDS-R score and the synthesis of several selected prosodic features. Consequently, the adjusted coefficient of determination ($\overline{R}^2 = 0.57$) suggests that prosody-based speech sound analysis could potentially be used to detect cognitive impairment in elderly subjects.

1 INTRODUCTION

Japan has a rapidly aging society and in 2005 had 2.05 million elderly patients with dementia. The number of the patients with dementia is expected to increase to more than 3 million over the next 10 years (Awata, 2009). Thus, the Ministry of Health, Labour and Welfare (MHLW) has begun projects to improve dementia treatment and quality of life. These projects are focused on the development of early detection methods for dementia that are both sensitive and specific.

To screen for dementia and cognitive impairment, a questionnaire test such as Mini-Mental State Examination (MMSE) (Folstein et al., 1975), Revised Hasegawa's Dementia Scale (HDS-R) (Imai and Hasegawa, 1994), Clinical Dementia Rating (CDR) (Morris, 1993), and Memory Impairment Screen (MIS) (Buschke et al., 1999), is commonly used in addition to a neurophysiological test (e.g., using MRI, FDG-PET, and CSF biomarkers). Questionnaire tests have some disadvantages and their use is limited in the clinic. The MMSE, HDS-R, and CDR are more

Kato S., Suzuki Y., Kobayashi A., Kojima T., Itoh H. and Homma A..
STATISTICAL ANALYSIS OF THE SIGNAL AND PROSODIC SIGN OF COGNITIVE IMPAIRMENT IN ELDERLY-SPEECH - A Preliminary Study.
DOI: 10.5220/0003189903220327
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2011), pages 322-327
ISBN: 978-989-8425-35-5

time-consuming that a general practitioner's consultation. In general, the questionnaire cannot completely dismiss the influence of education, social class, and gender difference on the results. In addition, there is a possibility that practitioner subjectivity may affect the scoring. Thus, we believe that the development of a simple, non-invasive examination that is objective and combined with a physiological test could enables the early detection of dementia in a broad population.

In a pilot study, we focused on speech sounds during the subject's answers to the questionnaire. Taler et al. have reported language (Taler and Phillips, 2007), and grammatical, and emotional prosodic impairment (Taler et al., 2008), as well as mild cognitive impairment (MCI), in elderly patients with Alzheimer's disease (AD). Hoyte et al. (Hoyte et al., 2009) reported that the components of speech prosody are useful for detecting the syntactic structure of speech. These reports suggest the possibility of using speech prosodic feature analysis to screen for dementia. This paper presents a novel approach to the early detection of cognitive impairment in the elderly that uses speech sound analysis in combination with a multivariate statistical technique. In this paper, we focused on the prosodic features of speech sound. We expect that the computation and information technology of this approach will enable general practitioners to easily screen for dementia. In our preliminary study, we examined the relationship between the HDS-R score and speech prosodic features. In addition, we addressed the effectiveness of speech prosody in discriminating among elderly individuals with normal cognitive abilities (NL), patients with mild cognitive impairment (MCI), and Alzheimer's disease (AD).

2 METHOD

2.1 Design

We recorded the speech sound of elderly patients while they provided answers for an HDS-R questionnaire test. We focused on questions questions about time orientation and number backward counting. In addition, we collected speech sounds while the patients were talking about the topics of hometown, childhood, and school.

2.2 Participants

Fifty six Japanese subjects (22 males and 34 females between the ages of 64 and 90 years) participated in this study. With some exception, we collected three samples of speech sound from each of the participants. The number of total sound data points was 146 as shown in Table 1. The sound data contained 42 samples of speech by elderly individuals with normal cognitive abilities (NL) (HDS-R score was 28.6 ± 1.7, n=16), 41 samples from patients with mild cognitive impairment (MCI) (HDS-R score was 25.4 ± 1.8, n=14), and 63 samples from patients with Alzheimer's disease (AD) (HDS-R score was 18.5 ± 5.5, n=22). Four subjects were excluded from these clinical groups because of the diffuse Lewy body disease (DLBD).

3 MEASUREMENT

3.1 Prosodic Feature Extraction

Speech has three components: prosody, tone, and phoneme. Past research indicates that the prosodic component has important non-verbal information such as emotional expressions (Cowie et al., 2001), (Scherer et al., 2003), (Cho et al., 2009). In accordance with our hypothesis, cognitive impairment was observed in the elderly (Taler and Phillips, 2007), (Taler et al., 2008). In this study, we considered 98 different acoustic correlates related to both segmental and suprasegmental information from speech signals. We used a computational data mining strategy based on a statistical-analytical approach. We extracted as many features as possible, and disregarded irrelevant features using a feature selection technique. These features were phrase-level statistics corresponding to fundamental frequency (F0) and their time-series behavior (13 features), formant and its time-series behavior (61 features), power envelope and its time-series behavior (22 features), speech rate, and response time (2 features). Prosodic analysis was performed in 23-ms frames and passed through a Hamming window (1024 points). Voice waveforms (sampled at 44.1 kHz with 16 bits) were extracted using a short-time Fourier transform (STFT) every 11 ms.

3.1.1 Spectral and Pitch Features

The set of 13 spectral features is comprised of statistical properties and time-series behaviors of fundamental frequency (F0).

F1.-7. Amplitude of F0 contour during t sec after the beginning of the phrase ($t = 0.05, 0.10, \cdots, 0.35$). The F0 contour is recorded in the interquartile range.

F8. Spectral centroid.

F9.-12. Standard deviation, mean, maximum, and minimum value of the F0 contour.

Table 1: Category Breakdown of the Speech Data (N=52).

Age	64-70	71-75	76-80	81-85	86-90	Total
Male	9 (6)	0 (0)	14 (5)	26 (9)	5 (2)	54 (19)
Female	16 (6)	18 (7)	23 (8)	24 (8)	11 (4)	92 (33)
Subtotal	25 (9)	18 (7)	37 (13)	50 (16)	16 (6)	146 (52)

Value in bracket means the number of subjects.

F13. Gradient of the linear regression line of the F0 contour.

3.1.2 Formant Features

Formant features consist of 61 values of frequency and bandwidth for the first 4 formants of distinguishing or meaningful frequency components within human speech.

For1. Power ratio of the sum of four formant components to whole frequency range.

For2.-4. Power ratio of the sum for the first n formant components to that of four formant components ($n = 1,2,3$).

For5. Power ratio between odd and even formant components.

For6.-9. Standard deviation of the first, second, third, and fourth formant frequencies.

For10.-13. Mean value of the first, second, third, and fourth formant frequencies.

For14.-17. Maximum value of the first, second, third, and fourth formant frequencies.

For18.-21. Minimum value of the first, second, third, and fourth formant frequencies.

For22.-25. Median value of the first, second, third, and fourth formant frequencies.

For26.-29. Difference between the maximum and minimum values of the first, second, third, and fourth formant frequencies.

For30.-33. Gradient of the linear regression line of the first, second, third, and fourth formant frequencies.

For34.-37. Standard deviation of the first, second, third, and fourth formant bandwidths.

For38.-41. Mean value of the first, second, third, and fourth formant bandwidths.

For42.-45. Maximum value of the first, second, third, and fourth formant bandwidths.

For46.-49. Minimum value of the first, second, third, and fourth formant bandwidths.

For50.-53. Median value of the first, second, third, and fourth formant bandwidths.

For54.-57. Difference between the maximum and minimum values of the first, second, third, and fourth formant bandwidths.

For58.-61. Gradient of the linear regression line of the first, second, third, and fourth formant bandwidths.

3.1.3 Intensity (Energy) Features

We extracted 22 energy features with the statistical properties of the power envelope.

Pow1. Gradient of the linear regression line of the power envelope.

Pow2.-8. Median value of the first derivative of the power envelope during the t seconds after the beginning of the phrase ($t = 0.05, 0.10, \cdots, 0.35$).

Pow9.-15. Ratio of the power at t seconds after the beginning of the phrase to the maximum power ($t = 0.05, 0.10, \cdots, 0.35$).

Pow16.-19. Standard deviation, mean, maximum, and minimum value of the short-time power.

Pow20. Gradient of the linear regression line of the power envelope from the beginning of the phrase to the peak.

Pow21. Gradient of the linear regression line of the power envelope from the peak to the end of the phrase.

Pow22. Ratio between the time from the beginning of the phrase to the peak and the time from the peak to the end.

3.1.4 Speech Rate and Response Time

In addition, we measured two features concerning speech rate and response time to answer in the questionnaire.

T1. Average duration for a single mora.

T2. Time taken to respond to the questionnaire.

3.2 Automatic Feature Selection

In our strategy for feature extraction, all of the prosodic features described above may not be equally useful and important for discrimination among NL, MCI, and AD. This creates the need for systematic feature selection. In this study, we used the forward stepwise (FSW) method (Draper and Smith, 1998), which is the most popular form of feature selection in statistics and consists of a combination of the forward selection and backward elimination methods. FSW is a greedy algorithm that adds the best feature (or deletes the worst feature) during each round. We chose a model selection method based on the Akaike's information criterion (AIC) (Akaike, 1974), which is a measure of the goodness of fit of an estimated statistical model. Using this criterion if the FSW, we were able to develop an estimation accuracy model with high accuracy and avoid over-fitting to training data. The AIC is defined as:

$$\mathrm{AIC} = -2\ln L + 2k, \tag{1}$$

where k is the number of parameters in the estimated model, and L is the maximized value of the likelihood function for the estimated model. Under the assumption that the model errors are normally and independently distributed, this becomes (up to an additive constant, which depends only on n and not on the model):

$$\mathrm{AIC} = n \cdot \ln(\mathrm{RSS}/n) + 2k, \tag{2}$$

where n is the number of data points (sample size), and RSS is the residual sum of squares from the estimated model. In this study, the RSS was obtained by calculating the sum of the square error of the difference between the estimated and observed HDS-R scores. FSW selects the best subset of all features to minimize (locally) the AIC score.

When determining the model parameters using the maximum likelihood estimation, it is possible to increase the likelihood by adding additional parameters; however, this also may result in over-fitting of the data. This represents a tradeoff between precision and complexity in the model. In addition to Schwarz's BIC (Schwarz, 1978), the AIC resolves this problem by introducing a penalty term (corresponding to the second term in Eq. (2)) for the number of parameters in the model. This penalty discourages over fitting, but it should be avoided so that the feature may be effectively eliminated. In this paper, we introduce a pre-processing method that synthesizes prosodic features by principal component analysis (PCA) prior to feature selection. This method is a combination of principal component regression (Massy, 1965) and automatic feature selection.

In the following section, the correlation between HDS-R score and synthesis of selected prosodic features is described by experimental results of multiple regression analysis through four manners of feature selection: forward stepwise method with AIC (FSW-AIC), PCA pre-processed forward stepwise method with AIC (PCA-FSW-AIC), forced entry method without feature selection (FE) , and PCA pre-processed forced entry method without feature selection (PCA-FE).

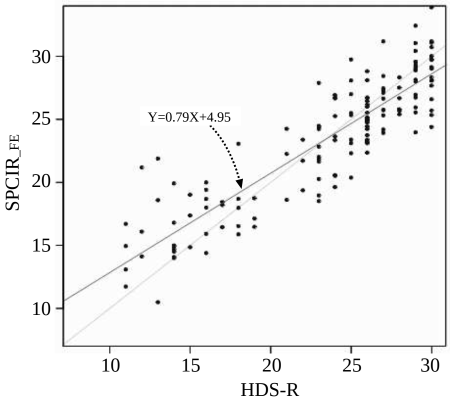

Figure 1: Scatter plot of HDS-R and SPCIR$_{\mathrm{FE}}$ ($\bar{R}^2 = 0.36$).

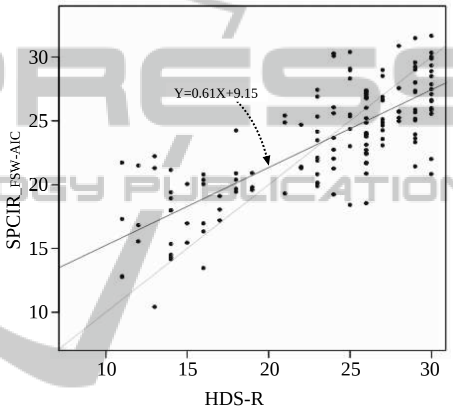

Figure 2: Scatter plot of HDS-R and SPCIR$_{\mathrm{FSW\text{-}AIC}}$ ($\bar{R}^2 = 0.50$).

4 RESULTS AND DISCUSSION

This section describes the correlation between HDS-R and speech prosody in elderly individuals using 146 speech voice samples (N=52), each with 98 prosodic features. We calculated the speech prosody-based cognitive impairment rating (SPCIR) by multiple linear regression using prosodic features (as regressors) selected by the feature selection method mentioned above. In PCA pre-processing, we used kernel PCA (Schölkopf et al., 1998) as the principal component analysis. and calculated 98 PCs because the accumulated contribution relevance is more than 90%.

SPCIR$_{\mathrm{FE}}$, SPCIR$_{\mathrm{FSW\text{-}AIC}}$, SPCIR$_{\mathrm{PCA\text{-}FE}}$, and SPCIR$_{\mathrm{PCA\text{-}FSW\text{-}AIC}}$ were calculated from the feature set chosen by FE, FSW-AIC, PCA-FE, and PCA-FSW-AIC, respectively. Table 2 shows the results of the analysis and the scatter plots of HDS-R and the SPCIRs are shown in Fig. 1-4. Table 3 shows the dominant regressors obtained from each of the feature selection methods.

Table 2: Correlation between SPCIR and HDS-R by Multiple Linear Regression.

	SPCIR_{FE}	$\text{SPCIR}_{\text{FSW-AIC}}$	$\text{SPCIR}_{\text{PCA-FE}}$	$\text{SPCIR}_{\text{PCA-FSW-AIC}}$
# of regressors	98	31	93	55
R	0.78	0.61	0.77	0.73
$\bar{R}^2$	0.36	0.50	0.35	0.57
S.E.	4.56	4.02	4.62	3.75

Table 3: Dominant Regressors for Estimate of HDS-R.

Method	dominant regressors	
SPCIR_{FE}		98 regressors in total
**	Pow8, For48	
*	For49, For5, For55, For51	
$\text{SPCIR}_{\text{FSW-AIC}}$		31 regressors in total
***	Pow7, Pow15, For48, For26, Pow18, F4, For47, For15, Pow9, Pow16	
**	Pow5, For29	
*	F13, For13, For46, For19, For50, For51, For21, For47	
$\text{SPCIR}_{\text{PCA-FE}}$		93 regressors in total
***	PC93	
**	PC7, PC34	
*	PC27,PC32, PC71, PC19, PC8, PC51	
$\text{SPCIR}_{\text{PCA-FSW-FE}}$		55 regressors in total
***	PC93, PC7, PC34	
**	PC29, PC32, PC71, PC19	
*	PC8, PC51, PC49, PC42, PC2, PC92, PC20, PC38, PC25, PC47, PC44, PC53, PC80, PC10, PC39, PC30	

***: with significance level of 0.001, **: with significance level of 0.01
*: with significance level of 0.05

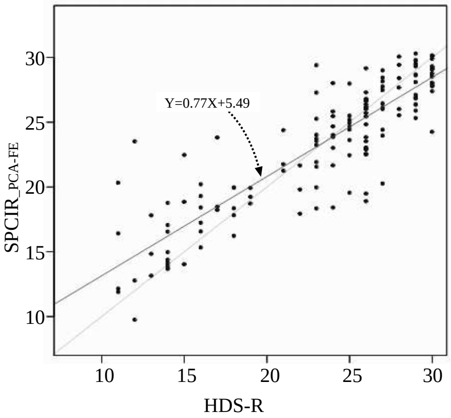

Figure 3: Scatter plot of HDS-R and $\text{SPCIR}_{\text{PCA-FE}}$ ($\bar{R}^2 = 0.35$).

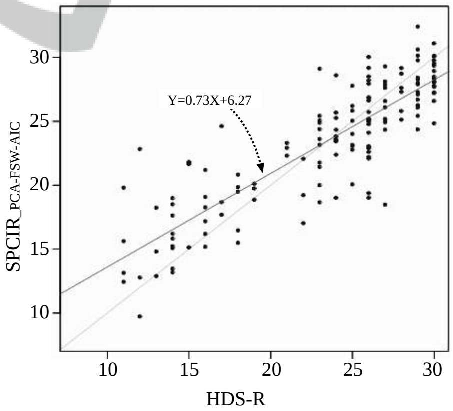

Figure 4: Scatter plot of HDS-R and $\text{SPCIR}_{\text{PCA-FSW-AIC}}$ ($\bar{R}^2 = 0.57$).

SPCIR_{FE} apparently has a larger correlation with HDS-R ($R = 0.78$); however the adjusted coefficient of determination declined ($\bar{R}^2 = 0.36$). This method detected few dominant regressors suggesting overfitting of the samples and multicollinearity due to a large number of regressors.

$\text{SPCIR}_{\text{FSW-AIC}}$ avoids the disadvantages of overfitting and increases the number of dominant regressors; however it does not give a satisfactory HDS-R correlation ($R = 0.61$) and an adjusted coefficient of determination ($\bar{R}^2 = 0.50$). $\text{SPCIR}_{\text{FSW-AIC}}$ uses only 31 total regressors due to the penalty term of AIC, which was based on model complexity. There might be effective features for estimation of HDS-R in the 67 regressors that were not chosen by FSW-AIC.

$\text{SPCIR}_{\text{PCA-FSW-AIC}}$, with the PCA pre-processed forward stepwise method in combination with AIC, solved the above-mentioned problems. In this method, principal components of 98 features were used as regressor candidates during feature selection, and 55 PCs were used as regressors in multiple regression. As shown in Table 3, the principal components with higher variance (i.e., PC7, PC8, PC2) were dominant regressors; however the low-variance principal components, such as PC93, PC71 and PC92, were also important for estimation of HDS-R. Finally, we obtained the scatter plot shown in Fig. 4, which suggests a positive linear relationship between HDS-R and SPCIR. The results indicates a moderately significant correlation ($R = 0.73$) between the HDS-R score and the appropriate synthesis of several selected prosodic features. Consequently, the adjusted coefficient of determination ($\bar{R}^2 = 0.57$) suggests that prosody-based speech sound analysis could potentially be used to detect cognitive impairment in elderly patients.

5 CONCLUSIONS AND FUTURE WORK

Our study presented a novel approach to detect cognitive impairment in elderly patients. This approach uses prosody-based speech sound analysis and a multivariate statistical technique. As a clinical data examination, we collected 146 speech voice samples from 56 Japanese participants and extracted 98 prosodic features from each of the samples. We then analyzed the correlation between the HDS-R score and synthesis of selected prosodic features by multiple linear regression in combination with sophisticated feature selection. We uncovered a moderately significant correlation. Thus, this speech prosody-based approach may be used to detect cognitive impairment in elderly patients. In future work, more expansive multimodality data collection will be performed using noninvasive neurophysiological measurements such as functional near-infrared spectroscopy (fNIRS). Much more clinical trials will also be evaluated, and the technique proposed here will be used as a screening tool for dementia.

ACKNOWLEDGEMENTS

We are grateful to Dr. H. Endo and National Center for Geriatrics and Gerontology for clinical assessment and speech data collection. This work was supported in part by SENTAN, Japan Science and Technology Agency (JST), and part by Suzuken Memorial Foundation.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.

Awata, S. (2009). Roll of the dementia medical center in the community. In *Japanese Journal of Geriatrics*, volume 46, pages 203–206. (in Japanese).

Buschke, H., Kuslansky, G., Katz, M., Stewart, W. F., Sliwinski, M. J., Eckholdt, H. M., and Lipton, R. B. (1999). Screening for dementia with the Memory Impairment Screen. *Neurology*, 52(2):231–238.

Cho, J., Kato, S., and Itoh, H. (2009). Comparison of Sensibilities of Japanese and Koreans in Recognizing Emotions from Speech by using Bayesian Networks. In *IEEE International Conference on Systems, Man, and Cybernetics*, pages 2945–2950.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. G. (2001). Emotion recognition in human-computer interaction. *IEEE Signal Processing Magazine*, 18(1):32–80.

Draper, N. and Smith, H. (1998). *Applied Regression Analysis (3rd edition)*. John Wiley & Sons.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). “Mini-Mental State”: A practical method for grading the cognitive state of patients for the clinician. *J. Psychiat. Res*, 12(3):189–198.

Hoyte, K., Brownell, H., and Wingfield, A. (2009). Components of Speech Prosody and their Use in Detection of Syntactic Structure by Older Adults. *Experimental Aging Research*, 35(1):129–151.

Imai, Y. and Hasegawa, K. (1994). The revised Hasegawa’s Dementia Scale (HDS-R): evaluation of its usefulness as a screening test for dementia. *J. Hong Kong Coll. Psychiatr.*, 4(SP2):20–24.

Massy, W. F. (1965). Principal Components Regression in Exploratory Statistical Research. *Journal of the American Statistical Association*, 60(309):234–256.

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology*, 43(11):2412–2414.

Scherer, K. R., Johnstone, T., and Klasmeyer, G. (2003). *Vocal expression of emotion*. R. J. Davidson, H. Goldsmith, K. R. Scherer eds., Handbook of the Affective Sciences (pp. 433–456), Oxford University Press.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319.

Schwarz, G. E. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Taler, V., Baum, S. R., Chertkow, H., and Saumier, D. (2008). Comprehension of grammatical and emotional prosody is impaired in Alzheimer’s disease. *Neuropsychology*, 22(2):188–195.

Taler, V. and Phillips, N. (2007). Language performance in Alzheimer’s disease and mild cognitive impairment: A comparative review. *Journal of Clinical and Experimental Neuropsychology*, 30(5):501–556.