Phase-Based Feature Representations for Improving Recognition of Dysarthric Speech

Conference Paper · December 2018

DOI: 10.1109/SLT.2018.8639031

3 authors:

Siddharth Sehgal
The University of Sheffield
8 PUBLICATIONS **103** CITATIONS
SEE PROFILE

Stuart Cunningham
The University of Sheffield
47 PUBLICATIONS **1,276** CITATIONS
SEE PROFILE

Phil Green
The University of Sheffield
128 PUBLICATIONS **4,089** CITATIONS
SEE PROFILE

PHASE-BASED FEATURE REPRESENTATIONS FOR IMPROVING RECOGNITION OF DYSARTHRIC SPEECH

Siddharth Sehgal[1], *Stuart Cunningham*[1], *Phil Green*[2]

[1]Department of Human Communication Sciences, University of Sheffield, UK
[2]Department of Computer Science, University of Sheffield, UK

ABSTRACT

Dysarthria is a neurological speech impairment, which usually results in the loss of motor speech control due to muscular atrophy and incoordination of the articulators. As a result the speech becomes less intelligible and difficult to model by machine learning algorithms due to inconsistencies in the acoustic signal and data sparseness. This paper presents phase-based feature representations for dysarthric speech that are exploited in the group delay spectrum. Such representations are found to be better suited to characterising the resonances of the vocal tract, exhibit better phone discrimination capabilities in dysarthric signals and consequently improve ASR performance. All the experiments were conducted using the UASPEECH corpus and significant ASR gains are reported using phase-based cepstral features in comparison to the standard MFCCs irrespective of the severity of the condition.

Index Terms— Dysarthric speech recognition, adaptation, group delay spectrum, phase-based cepstrals

1. INTRODUCTION

Dysarthria is the collective name for a group of neurological speech disorders which result from damage to the central or peripheral nervous system. Dysarthric speech is usually characterised by symptoms such as reduced stress, slow speech rate, hypernasality, muscular rigidity, spasticity, monopitch and limited range of speech movements [1, 2]. It can have debilitating effects on speech production and can simultaneously affect subglottal, laryngeal and articulatory movements [3]. The most prevelant causes of such motor speech disorder in the UK are stroke, cerebral palsy and Parkinson's disease [4]. Reports suggest that there is an ever-growing need to improve human-to-machine interaction for people with dysarthria in order to promote overall wellbeing and independence [5]. People with dysarthria are often physically disabled, so speech provides an attractive interface for a natural and faster mode of interaction [6, 7]. It has been shown in the past that such an interface can naturally be extended to control electronic devices and as a communication aid [7, 8].

Some of the difficulties in recognising dysarthric speech are the high degree of inter and intra-speaker variations, data sparsity issues and malformed phonetic space. Broadly categorising, researchers have tried to address these problems in three ways: (i) Acoustic modelling using both generative and discriminative techniques, (ii) Speaker adaptation approaches and (iii) Signal transformation and enhancement techniques.

Recent studies have shown that speaker-adapted (SA) systems are generally better at modelling dysarthric variabilities, regardless of the severity [9, 10, 11]. SA systems usually require a good initial speaker-independent (SI) model to begin with, and it has been shown that a systematic approach to selecting the initial model can boost recognition performance [12, 13, 14]. It has also been shown that in addition to typical adaptation using MLLR and MAP, hybrid adaptation approaches are preferable for building SA models and speaker adaptive training (SAT) further improves overall performance by implicitly handling the inter-speaker variabilities [11].

Although the majority of work in ASR on dysarthric speech has been done using the HMM-GMM framework, other hybrid architectures like DNN-HMM [15, 16, 17] and HMM-SVM [18] are slowly gaining prominence. The application of these recent techniques is reported on only constrained small vocabulary tasks with minimal performance improvement. However, in some comparable studies the reported results using the DNN-HMM framework [19] were not able to outperform the standard HMM-GMM systems that deployed hybrid adaptation techniques [11].

In addition to the acoustic modelling and adaptation techniques, speaker-specific pronunciation adaptation has also been successfully expoited to improve performance [12, 20, 21]. Lastly, signal enhancement techniques have been pursued for dysarthric classification, intelligibility and ASR improvements. Indirect approaches have looked at suggesting informative frequency bands [22] and finding optimal MFCC configurations [23], whereas other approaches perform direct signal manipulation [24, 25, 26] or produce enriched MFCCs with perceptually motivated features [27, 16].

This paper focuses on exploiting alternative feature representations for dysarthric speech that are based on the phase spectrum of the Fourier analysis instead of the standard magnitude spectrum. Our aim is to show that cepstral features based on the phase information in the signal are better at characterising dysarthric speech in comparison to magnitude

based features, which consequently improve the ASR performance. The paper is organised as follows: Section 2 will give a theoretical overview of the phase feature representations and its benefits for characterising dysarthric speech, Section 3 details the experimental setup and the ASR methodology used, Section 4 shows the empirical work comparing the ASR results for the phase-based and standard MFCC feature sets, Section 5 discusses the outcome of the ASR experiments from a statistical viewpoint and Section 6 concludes.

2. PHASE FEATURE REPRESENTATIONS

Fourier analysis breaks the speech signal into its fundamental constituents and encodes information for the observed frequencies in its respective magnitude and phase components. Despite the fact that both magnitude and phase are needed for the true representation of any speech signal, most of the conventional feature representations for ASR only exploit the magnitude spectrum and the phase spectrum is mostly ignored. The historical reason for disregarding the phase spectrum is the inability of the human ear to resolve phase information [28, 29]. Another difficulty is the chaotic nature of the phase spectrum which results from random polarity and wrapping into the range $\pm\pi$.

One of the earliest references showing the importance of phase was detailed in a systematic study conducted by Oppenheim and Lim [30]. Their paper highlighted the prominence of phase-only synthesis in analysing atomic crystal structures for measuring contours of electron density. It further gave illustrative examples where phase-only reconstruction of images and speech signals were closer to the original than magnitude-only reconstruction. It also showed that phase-only reconstruction of speech signals was better at preserving key "event locations" and had better correlation to the original signal.

The remainder of this section will outline some key representations of speech signals that are based on the processing of the phase spectrum. These phase-based features will be used in later sections, which investigate their efficacy in characterising and recognising dysarthric speech in comparison to standard magnitude-based MFCCs.

2.1. Group delay function

The processing of the phase spectrum is a difficult task due to wrapping constraints of the spectrum between the values of $\pm\pi$. The wrapping causes the phase spectrum to be a chaotic curve with random fluctuations. In order to overcome this problem, the phase spectrum can be computed as the negative first order derivative of the unwrapped phase spectrum. This is known as the **Group Delay Function (GDF)** and is mathematically expressed as:

$$\tau(\omega) = -\frac{d(\phi(\omega))}{d\omega} \tag{1}$$

The above equation represents the *"rate of change in the phase spectra"*, where $\phi(\omega)$ is the continuous unwrapped phase spectrum. The unwrapping of phase involves adding multiples of $\pm 2\pi$ whenever the alignment between consecutive frequency bins exceeds π. The unwrapping process is not always straightforward and it can also be completely avoided by computing the phase spectrum from the time-domain signal [31] as:

$$\tau(\omega) = \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{|X(\omega)|^2} \tag{2}$$

where $X(\omega)$ and $Y(\omega)$ are the Fourier transforms of the discrete-time real signals $\{x(n)\}$ and $\{nx(n)\}$ respectively and R and I denote the real and imaginary parts of the complex output.

The motivation behind representing the speech signal using the phase spectrum instead of the more commonly used magnitude spectrum lies in the properties of the GDF. The theory, properties and its practical applications are discussed in greater detail in [32, 33]. However, the two most important properties of the GDF are the:

(I) Property of Additivity: The convolution of any time-domain signal is additive in the group delay phase spectra. This contrasts with the magnitude spectra, which are multiplicative.

(II) Property of Higher Resolution: The GDF tends to exhibit higher resolving power in differentiating closely spaced resonance peaks in the spectrum.

Despite the advantages of the group delay spectrum, it comes with a caveat, which can be detrimental for front-end processing in ASR. If there are zeros which occur too close to the unit circle (highly likely in a speech signal due to its mixed-phase nature), these can result in huge spikes in the group delay spectrum. The spikes tend to dominate the spectral shape and obscure the true locations of the formants making the spectrum not very useful for feature generation. The occurrence of spikes in the spectrum results when the denominator term $|X(\omega)|^2$ in equation 2 gets smaller, i.e., when the distance between the zero location and the corresponding frequency bin on the unit circle reduces.

2.2. Phase-based features for dysarthric speech

Any meaningful representation of phase features based on the GDF will directly aim at reducing the inadvertent spikes introduced by the smaller values of $|X(\omega)|^2$ in equation 2. One such representation is the modified group delay function (MODGDF) [34] that was formulated to reduce such detrimental effects in order to maintain the dynamic range of the spectrum. It was shown that by introducing $|S(\omega)|$, which is a cepstrally smoothed version of $|X(\omega)|$, very low values can be avoided in the denominator of equation 2. The modified group delay function is defined as:

$$\tau_{MODGDF}(\omega) = \left(\frac{\tau_X(\omega)}{|\tau_X(\omega)|}\right)\left(|\tau_X(\omega)|\right)^{\alpha} \quad (3)$$

$$\tau_X(\omega) = \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{|S(\omega)|^{2\gamma}} \quad (4)$$

where $S(\omega)$ is the cepstrally smoothed version [35] of $X(\omega)$. In addition, the parameters α,γ can be empirically controlled to reduce the effect of spikes in the modified group delay function.

Another alternative form of the GDF is known as the product spectrum. It includes information from both the magnitude and phase spectrum. It is defined as the product of the GDF and the power spectrum (PS) [36] denoted as:

$$\begin{aligned}\tau_{PS}(\omega) &= |X(\omega)|^2\,\tau(\omega) \\ &= X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)\end{aligned} \quad (5)$$

As a consequence of the definition of the product spectrum, the denominator term of $|X(\omega)|^2$ that was responsible for the spikes in the group delay spectrum is cancelled out. This can be a useful representation, since it exploits the benefits of both the power spectrum and the phase spectrum without any need for applying smoothing techniques.

The MODGDF and PS representations of the group delay function are used in this paper to extract the new cepstral features. The smoothing parameters of MODGDF were determined empirically with optimal values set at $\alpha = 0.95$ and $\gamma = 0.20$ for the experiments. A 26-band mel spaced triangular filter was used for the filterbank analysis. It was also observed that certain window operations affected the group delay spectrum to a greater degree by the introduction of spurious spikes, whilst others produced a much smoother spectrum. The MODGDF- and PS-based spectra were processed using the Hanning-Poisson window that gave the best resolution. MODGDF- and PS-based cepstral coefficients will be referred to as **MODGDFCC** and **PSCC**.

The conceptual understanding of the phase spectrum along with its properties forms a convincing and sufficient basis to extend the idea of such features for representing dysarthric speech. To the best of our knowledge, there is no work in the literature that explores the possibility of phase features for characterising dysarthric speech and evaluating its performance on ASR systems. This section will conclude by showing illustrative examples of how the MODGDFCC/PSCC features compare to the standard MFCC for representing dysarthric speech.

2.2.1. High resolution impact

A comparison between the MFCC and group delay spectra is shown in figure 1 for a speaker chosen from each of the intelligibility groups. The comparison was conducted for the front high vowel /iy/ in the production of the word **be**. Around 1000 samples of voiced segment were selected from the centre of the vowel /iy/ in generating the spectra. It is observed that across all the intelligibility groups, the group delay spectrum shows high-resolution and prominent peaks in comparison to the standard spectrum. For example, the frequency regions between 3-6 KHz (very-low), 2-4 KHz (low) and 2-4/6-8 KHz (high) are very loosely resolved for the MFCC spectrum, whereas, the group delay spectrum shows a highly resolved delineation of closely spaced possible formants. Also, the frequencies regions around 5 KHz (low), 4-7 KHz (mid) shows highly prominent peaks for the group delay spectrum in comparison to the imperceptible peaks exhibited by the standard MFCC-based spectrum. Both spectra are generated from the MFCC and PSCC based cepstral features.

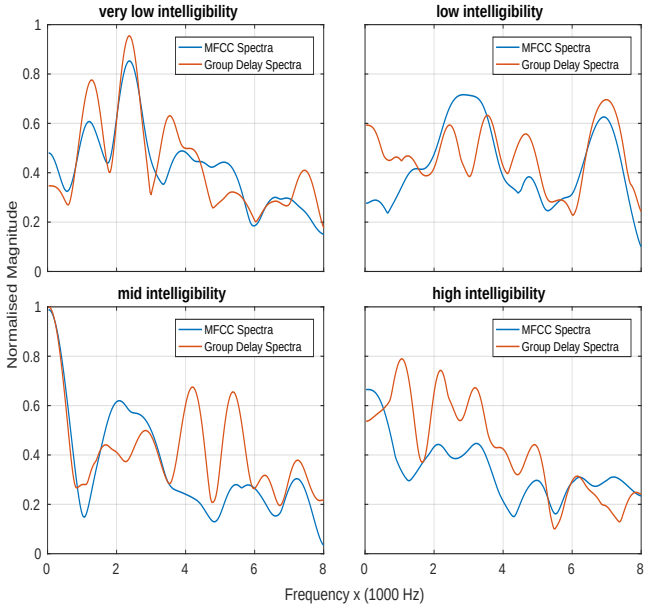

Fig. 1: Comparison of MFCC and group delay based spectral representations for a speaker chosen from each of the intelligibility groups in UASPEECH database. The comparison is for the vowel /iy/. Speakers selected for each intelligibility group are very-low→M04, low→F02, mid→M05, high→F05.

2.2.2. Better class separability

The aim of any feature representation technique for speech recognition is to capture sufficient discriminatory information about the individual phonetic tokens. The class (phonetic) separability is illustrated by examining the fricative sounds /s/ and /sh/ in context of the following vowel /iy/ in the words **see** and **she**. MFCC and PSCC features are compared for representing the example fricative syllable. It is emphasised that only PSCC representation will be shown for brevity: MODGDFCC gives similar results. Since the cepstral representa-

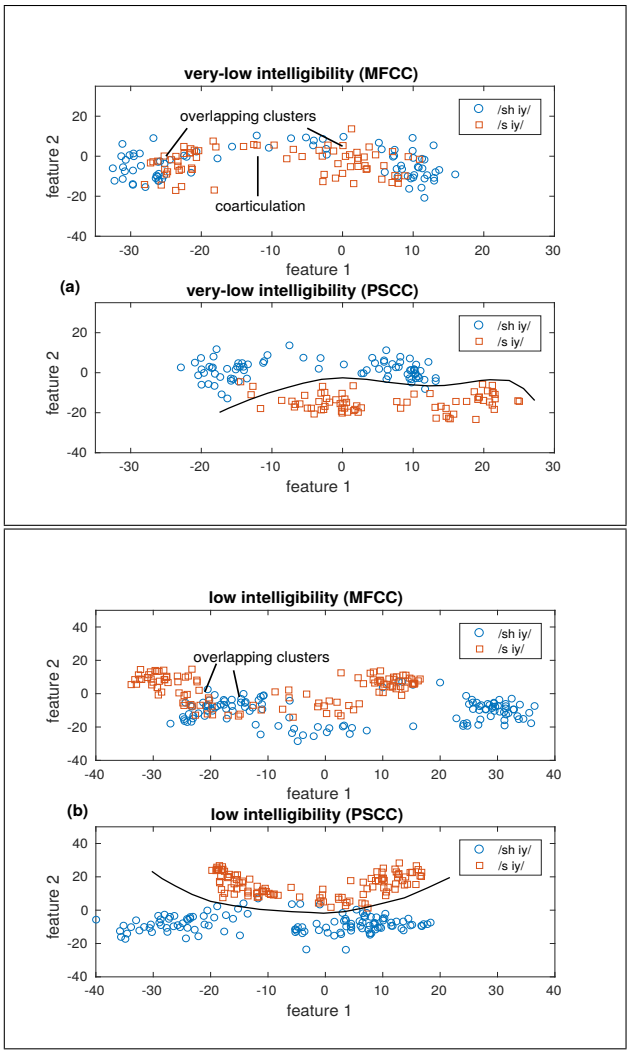

Fig. 2: Two dimensional projection for the MFCC & PSCC features for very-low and low intelligibility speakers.

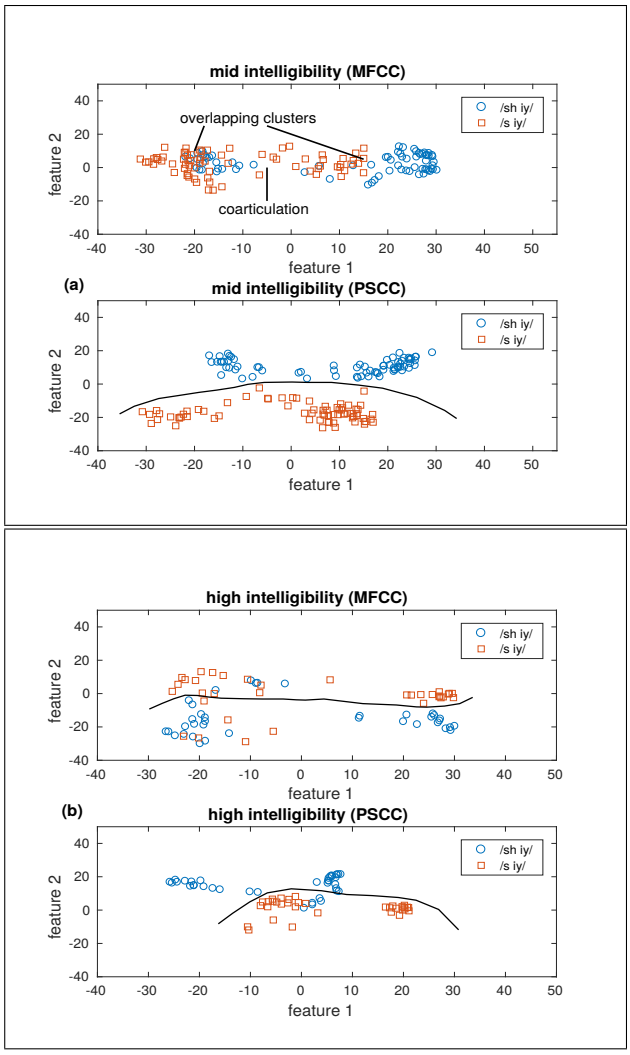

Fig. 3: Two dimensional projection for the MFCC & PSCC features for mid and high intelligibility speakers.

tion is generally encoded in higher dimensional space, Principal Component Analysis (PCA) is used for data visualisation in a lower dimensional subspace.

Figure 2 and Figure 3 show the two dimensional projection of the MFCC and PSCC features for the two syllable fricatives of a dysarthric speaker from each of the intelligibility groups. The MFCC projection clearly exhibits overlapping clusters in the very-low, low and mid intelligibility speakers, and the effects of coarticulation can be seen for the very-low and mid intelligibility speakers. In contrast, the PSCC representation for the same speech shows much better discriminatory capabilities. It shows the presence of more tightly bound clusters, which are easily separable in the acoustic space, and the coarticulatory effect of the two syllable fricatives tend to be non-overlapping.

Although the difference between MFCC and PSCC representation seems to reduce for the high intelligibility group, it can still be observed that the PSCC clusters for /s/ and /sh/ exhibit non-overlapping clusters in contrast to MFCC that still shows a noticeable overlap between the cloud of distinct points. Hence, the phase-based spectrum might prove better for phoneme discrimination of speech.

3. SETUP

3.1. Data preparation and modelling

The study used three databases for all the analysis and experiments. The Wall Street Journal corpus (WSJ SI-84) [37] and its British equivalent (WSJCAM0) [38] were used as typical speech data. In addition, the UASPEECH [39] corpus supplied dysarthric data. UASPEECH consists of data from 15 dysarthric and 13 control speakers. There are 765 isolated word occurances per speaker (455 distinct) collected

in three separate blocks (B1, B2, B3), where each block consists of 10 digits, 26 international radio words, 19 computer commands, 100 common words and 100 distinct uncommon words, which were not repeated across blocks. All the B1+B3 data is used for training and adaptation purposes and B2 is used for all the results reported in the paper. All the data was processed as standard magnitude-based MFCCs and phase-based MODGDFCC and PSCC features. Each of the cepstral features was represented as a 39-dimensional vector with 12 static coefficients, $c0$, first and second order time derivatives. For acoustic modelling, continuous density HMMs were used, which were word internal tied-state triphone models with a strict left-to-right topology. Clustering was performed using phonetic decision trees. Finally the number of Gaussian components was increased to 16 per state and the silence states were modelled using 32 Gaussian components.

3.2. Methodology

All the ASR experiments used a subset of speech systems from our earlier study in [11] and the relative performance is compared to our previous results. To the best of our knowledge, the results in [11] are the best reported on this relatively large database with a vocabulary of 255 distinct words. Table 1 summarises the speech systems tested. SI & SAT models were adapted using the hybrid MLLR-MAP approach.

System	Training Dataset Used
SD	UASPEECH-Dysarthria
SI-00	WSJ SI-84 + WSJCAM0
SI-02	UASPEECH-Dysarthria
SAT	UASPEECH-Dysarthria

Table 1: ASR systems and the corpora used for the preparation.

4. EXPERIMENTS

The experiments are conducted to compare the performance between the standard MFCC and phase based PSCC & MODGDFCC representations. The results are shown in figure 4, where the x-axis is representative of the standard MFCC results presented in our earlier work in [11]. The plots show the relative ASR gains across the four tested speech systems (SD, SI-00 (mllrmap), SI-02 (mllrmap), SAT (mllrmap)). It is evident that the phase-based feature representations of dysarthric speech show gains across all the tested systems.

5. DISCUSSION

Both PSCC and MODGDFCC features were highly effective for modelling dysarthric speech with a greater degree of pathological disorder in comparison to standard magnitude

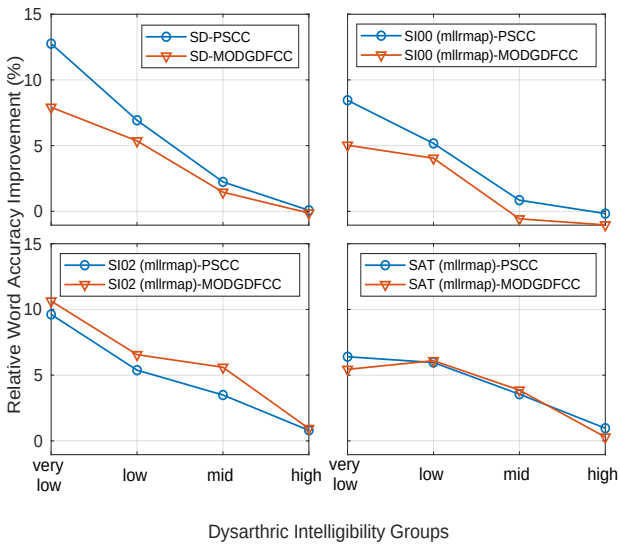

Fig. 4: Relative Word Accuracy gains for the PSCC and MODGDFCC based cepstral coefficients.

based MFCC. The relative benefit to ASR performance is reduced for speakers with less severe disorder. In order to statistically investigate the ASR performance gains of phase-based features, a pairwise Cochran's Q test was conducted for MODGDFCC/Standard-MFCC and PSCC/Standard-MFCC feature representations. Table 2 shows the absolute ASR word accuracy scores for the two feature representations. The cells that exhibit significant gains are marked with a †† ($p < 0.01$) or † ($p < 0.05$).

Out of the 16 possible combinations between the four systems and intelligibility groups, PSCC features show significant gains in 13 systems and MODGDFCC features show significant gains in 12 systems. It is noteworthy that for both the feature representations, all the systems showed highly significant gains for the *very-low* and *low* intelligibility groups. This is an encouraging outcome, since the majority of dysarthric speech systems are primarily targeted to benefit users with a high degree of speech disorder. Hence, feature representation based on group delay spectra of pathological speech could prove to be significantly beneficial for robust acoustic modelling.

It was also noted that PSCC was significantly better (†) for speaker-dependent modelling over MODGDFCC for speakers with lowest intelligibility. The selection between PSCC and MODGDFCC seems to be a matter of choice and can be dependent on particular applications. MODGDFCC also comes with an additional constraint of finding optimal values of (α, γ), which can be dependent on the underlying dataset, whereas PSCC is free from such constraints and can benefit from the information in both the magnitude and phase spectrum.

In future work it would be interesting to explore if ASR

Intelligibility	PSCC Features				MODGDFCC Features			
	SD	**SI-02**	**SAT**	**SI-00**	**SD**	**SI-02**	**SAT**	**SI-00**
very-low	23.52	27.36	28.71	20.61	23.52	27.36	28.71	20.61
	26.52 ††	30.00 ††	30.55 ††	22.36 ††	25.33 ††	30.27 ††	30.28 ††	21.65 †
low	62.48	62.92	62.98	57.89	62.48	62.92	62.98	57.89
	66.81 ††	66.30 ††	66.72 ††	60.89 ††	65.82 ††	67.05 ††	66.83 ††	60.23 ††
mid	64.08	68.51	69.54	66.12	64.08	68.51	69.54	66.12
	65.52 †	70.90 ††	72.02 ††	66.69	65.02	72.34 ††	72.23 ††	66.23
high	83.07	86.17	86.87	87.08	83.07	86.17	86.87	87.08
	83.14	86.86 †	87.71 ††	86.93	82.96	86.98 †	87.12	86.19

Table 2: Absolute ASR word accuracy averaged by various intelligibility groups. The top number in each cell represents the best baseline results presented in [11] using standard MFCC features. The shaded number is the result of using phase-based feature representation for the MFCCs. Significant statistical gains are shown using a † ($p < 0.05$) or †† ($p < 0.01$).

of dysarthric speech could further benefit from both magnitude and phase-based cepstral representations by examining the efficacy of extended feature sets that will use information in both MFCC and PSCC/MODGDFCC.

6. CONCLUSION

Phase features extracted from the group delay spectrum were exploited for dysarthric speech signals. The cepstral features were extracted using the modified group delay (MODGDFCC) and product spectrum (PSCC) functions. Relative to standard MFCCs, the phase-based features were found to be better suited to characterising the resonances of the vocal tract, and exhibited better phone discrimination capabilities in dysarthric speech signals. Phase based cepstral features also showed significant gains in nearly all the tested speech systems and intelligibility groups. The phase features were most effective for the speakers with lowest intelligibility, with greatest relative gains. The results in the paper are based on a relatively large vocabulary of 255 words, whereas dysarthric speech systems usually operate over a much more constrained vocabulary set. Hence, in the light of these results it seems plausible to evaluate this approach in real applications in the future.

7. REFERENCES

[1] F.L. Darley, A.E. Aronson, and J.R. Brown, "Clusters of deviant speech dimensions in the dysarthrias," *Journal of Speech and Hearing Research*, vol. 12, pp. 462–496, 1969.

[2] J.R. Duffy, *Motor Speech Disorders : Substrates, Differential Diagnosis, and Management*, Elsevier Mosby, second edition, 2005.

[3] R.D. Kent, J.F. Kent, G. Weismer, and J.R. Duffy, "What dysarthria can tell us about the neural control of speech," *Journal of Phonetics*, vol. 28, no. 3, pp. 273–302, 2000.

[4] RCSLT, *Communicating Quality 3: RCSLT's Guidance on Best Practice in Service Organisation and Provision*, Royal College of Speech & Language Therapists, 2006.

[5] "Communication matters research matters: an aac evidence base," https://tinyurl.com/y7c6km5b, 2013, Online; accessed on: 08-Mar-2017.

[6] M. S. Hawley, "Speech recognition as an input to electronic assistive technology," *The British Journal Of Occupational Therapy*, vol. 65, no. 1, pp. 15–20, 2002.

[7] M. S. Hawley, P. Enderby, P. Green, S. Cunningham, S. Brownsell, J. Carmichael, M. Parker, A. Hatzis, P. O'Neill, and R. Palmer, "A speech-controlled environmental control system for people with severe dysarthria.," *Med Eng Phys*, vol. 29, no. 5, pp. 586–593, 2007.

[8] M. Hawley, S. Cunningham, P. Green, P. Enderby, R. Palmer, S. Sehgal, and P. O'Neill, "A voice-input voice-output communication aid for people with severe speech impairment," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 1, pp. 23–31, 2013.

[9] H.V. Sharma and M. Hasegawa-Johnson, "State-transition interpolation and map adaptation for hmm-based dysarthric speech recognition," in *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, 2010, pp. 72–79.

[10] H. Christensen, S. Cunningham, C. Fox, P. Green, and T. Hain, "A comparative study of adaptive, automatic recognition of disordered speech," in *13th Annual Conference of the International Speech Communication Association 2012, INTERSPEECH 2012*, 2012, vol. 2, pp. 1774–1777.

[11] S. Sehgal and S. Cunningham, "Model adaptation and adaptive training for the recognition of dysarthric

speech," in *6th Workshop on Speech and Language Processing for Assistive Technologies*, 2015.

[12] K.T. Mengistu and F. Rudzicz, "Adapting acoustic and lexical models to dysarthric speech," 2011, pp. 4924–4927.

[13] M.J. Kim, J. Yoo, and H. Kim, "Dysarthric speech recognition using dysarthria-severity-dependent and speaker-adaptive models," in *Interspeech*, 2013, pp. 3622–3626.

[14] H.V. Sharma and M. Hasegawa-Johnson, "Acoustic model adaptation using in-domain background models for dysarthric speech recognition," *Computer Speech and Language*, vol. 27, no. 6, pp. 1147–1162, 2013.

[15] C. Espaa-Bonet and J.A.R. Fonollosa, "Automatic speech recognition with deep neural networks for impaired speech," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10077 LNAI, pp. 97–107, 2016.

[16] C. Bhat, B. Vachhani, and S. Kopparapu, "Recognition of dysarthric speech using voice parameters for speaker adaptation and multi-taper spectral estimation," in *Interspeech*, 2016, pp. 228–232.

[17] N.M. Joy and S. Umesh, "Improving acoustic models in torgo dysarthric speech database," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2018.

[18] S. Chandrakala and N. Rajeswari, "Representation learning based speech assistive system for persons with dysarthria," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 9, pp. 1510–1517, 2017.

[19] S. Tejaswi and S. Umesh, "Dnn acoustic models for dysarthric speech," in *2017 23rd National Conference on Communications, NCC 2017*, 2017.

[20] S.O.C Morales and S.J. Cox, "Modelling errors in automatic speech recognition for dysarthric speakers," *EURASIP J. Adv. Signal Process*, pp. 2:1–2:14, 2009.

[21] W.K. Seong, J.H. Park, and H.K. Kim, "Dysarthric speech recognition error correction using weighted finite state transducers based on context-dependent pronunciation variation," in *Proceedings of the 13th international conference on Computers Helping People with Special Needs - Volume Part II*, 2012, vol. 7383, pp. 475–482.

[22] P.D. Polur and G.E. Miller, "Effect of high-frequency spectral components in computer recognition of dysarthric speech based on a mel-cepstral stochastic model," *Journal of Rehabilitation Research and Development*, vol. 42, no. 3, pp. 363–371, 2005.

[23] S.R. Shahamiri and S.S.B. Salim, "Artificial neural networks as speech recognisers for dysarthric speech: Identifying the best-performing set of mfcc parameters and studying a speaker-independent approach," *Advanced Engineering Informatics*, vol. 28, no. 1, pp. 102–110, 2014.

[24] H. Tolba and A.S. El Torgoman, "Towards the improvement of automatic recognition of dysarthric speech," in *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*, 2009, pp. 277–281.

[25] F. Rudzicz, "Adjusting dysarthric speech signals to be more intelligible," *Computer Speech & Language*, vol. 27, no. 6, pp. 1163–1177, 2013.

[26] C. Bhat, B. Vachhani, and S. Kopparapu, *Improving Recognition of Dysarthric Speech Using Severity Based Tempo Adaptation*, pp. 370–377, Springer International Publishing, 2016.

[27] S.-A. Selouani, H. Dahmani, R. Amami, and H. Hamam, "Using speech rhythm knowledge to improve dysarthric speech recognition," *International Journal of Speech Technology*, vol. 15, no. 1, pp. 57–64, 2012.

[28] G. S. Ohm, "Ueber die definition des tones, nebst daran geknpfter theorie der sirene und hnlicher tonbildender vorrichtungen," *Annalen der Physik*, 1843.

[29] H.L.F. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music.*, Longmans Green and Co., fourth edition, 1912.

[30] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *IEEE Proceedings*, vol. 69, pp. 529–541, 1981.

[31] A.V. Oppenheim and R.W. Schafer, *Discrete-time signal processing*, Prentice Hall, international edition, 1989.

[32] H.A. Murthy and B. Yegnanarayana, "Formant extraction from group delay function," *Speech Communication*, vol. 10, no. 3, pp. 209–221, 1991.

[33] H.A. Murthy and B. Yegnanarayana, "Group delay functions and its applications in speech technology," *Sadhana*, vol. 36, no. 5, pp. 745–782, Oct 2011.

[34] H.A. Murthy and V. Gadde, "The modified group delay function and its application to phoneme recognition," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, April 2003, vol. 1, pp. I–68–71.

[35] B. Yegnanarayana and H.A. Murthy, "Significance of group delay functions in spectrum estimation," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2281–2289, 1992.

[36] D. Zhu and K.K. Paliwal, "Product of power spectrum and group delay function for speech recognition," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2004, vol. 1, pp. I–125–8.

[37] D.B. Paul and J.M. Baker, "The Design for the Wall Street Journal-based CSR Corpus," in *Proceedings of the Workshop on Speech and Natural Language*, 1992, HLT '91, pp. 357–362.

[38] T. Robinson, J. Fransen, D. Pye, J. Foote, and S. Renals, "WSJCAMO: a British English speech corpus for large vocabulary continuous speech recognition," in *International Conference on Acoustics, Speech, and Signal Processing, ICASSP-95.*, 1995, vol. 1, pp. 81–84.

[39] H. Kim, M. Hasegawa-Johnson, A. Perlman, J. Gunderson, T.S. Huang, K. Watkin, and S. Frame, "Dysarthric speech database for universal access research," in *INTERSPEECH 2008, 9th Annual Conference of the International Speech Communication Association, Brisbane, Australia, September 22-26, 2008*, 2008, pp. 1741–1744.