On the Importance of Analytic Phase of Speech Signals in Spoken Language Recognition

Karthika Vijayan, Haizhou Li, Hanwu Sun, Kong Aik Lee

Department of Electrical and Computer Engineering, National University of Singapore
Human Language Technology Department, Institute for Infocomm Research, A*STAR, Singapore
Emails:{vijayan.karthika, haizhou.li}@nus.edu.sg, {hwsun, kalee}@i2r.a-star.edu.sg

Spoken Language Recognition

- To identify or verify the language identity in a speech segment
- Phonetic and phonotactic features differentiate languages/dialects
- Key interest in this paper is to explore the use of long-time information

Objective

To study the role of analytic phase of speech signals on human perception of spoken languages and automatic language recognition.

Analytic phase of speech

Long-time processing of speech

Multi-band demodulation analysis (MDA)

- Narrowband (NB) segmentation of speech
- Hilbert transform for each NB component, $s[n]$

$$S_H[k] = \begin{cases} 0, & k = 0, N/2 \\ S[k], & 1 \leq k \leq \frac{N}{2} - 1 \\ -S[k], & \frac{N}{2} + 1 \leq k \leq N - 1 \end{cases}$$

- Discrete-time analytic signal

$$z[n] = s[n] + js_H[n]$$

- Temporal amplitude and analytic phase
$a[n] = |z[n]|, \theta[n] = \angle z[n]$

Analytic phase

$$\theta[n] = tan^{-1}\left(\frac{s_H[n]}{s[n]}\right)$$

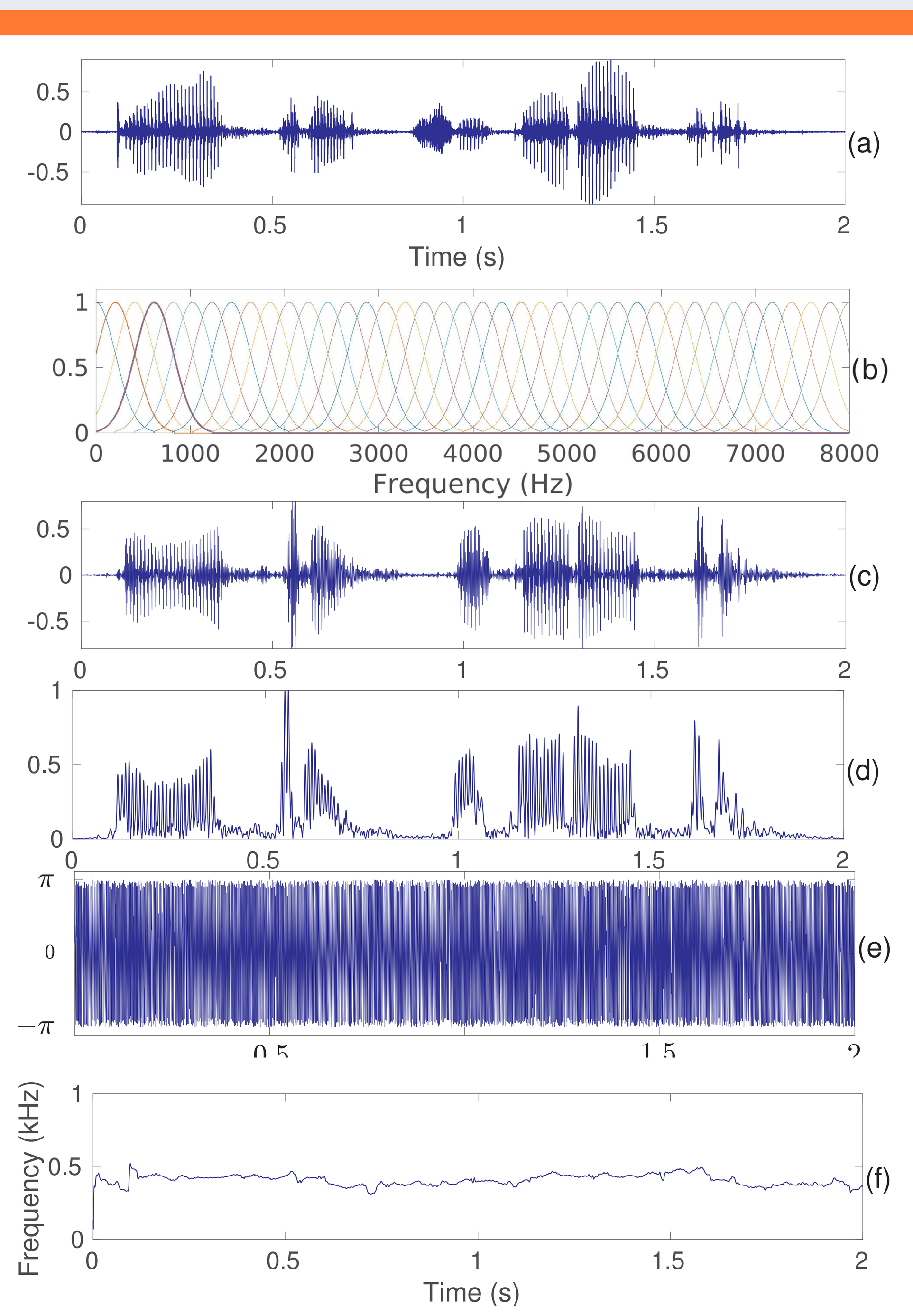

Figure 1: MDA: (a) Speech signal, (b) Gabor filter-bank, (c) NB component (d) Temporal amplitude $a[n]$, (e) Analytic phase $\theta[n]$ and (f) smoothed IF $\theta'[n]$.

Significance of analytic phase

Human perception of languages

- Speech Signals: Original vs Analytic phase-tampered
- Languages: Chinese Mandarin vs Chinese Min
  English-British vs English-American
- Analytic phase crucially affects perception of similar sounding languages

Table 1: Human language identification accuracy (%).

Type of speech	Chinese	English
Analytic phase-tampered	52	64
Original	94	96

Analytic phase for automatic SLR

Feature extraction

- Computation of analytic phase gets affected by phase wrapping
- Unambiguous representation of analytic phase by its derivative

Instantaneous frequency (IF)

$$\theta'[n] = \frac{2\pi}{N}\text{Re}\left\{\frac{\mathscr{F}^{-1}(kZ[k])}{\mathscr{F}^{-1}(Z[k])}\right\},$$

- Pyknogram - scatter plot of IFs computed from multiple NB components of speech
- Pyknogram shows the information captured by IF from speech signals

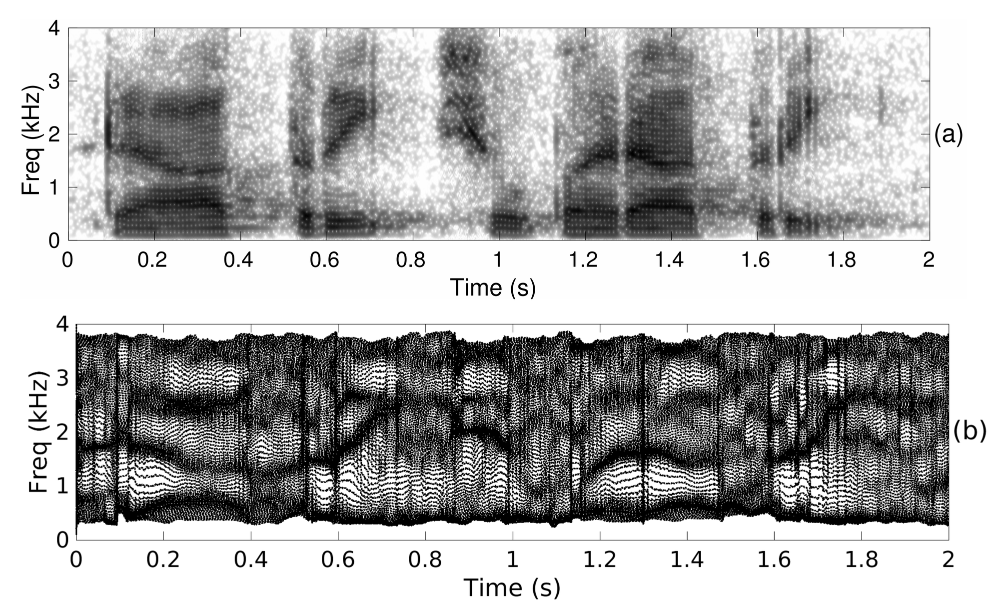

Figure 2: (a) Spectrogram and (b) Pyknogram corresponding to the segment of speech shown in Figure 1(a).

Features from analytic phase :-

- Long-time IF contours - segmented & averaged- IF coefficients (IFC)
- Discrete cosine transform of IFC- IF cepstral coefficients (IFCC)

Commonly used features in SLR

- Mel frequency cepstral coefficients (MFCC)
- Shifted delta cepstral coefficients (SDCC)
- Deep bottleneck features (DBN)
- SDCC and DBN capture long-time information in speech from spectral magnitude

IFCC vs other features

Feature	Window	Process	LT info	Source of LT info
MFCC	TD	ST	No	Nil
SDCC	TD	ST	Yes	Inter-frame relations
DBN	TD	ST	Yes	Inter-frame relations
IFCC	**FD**	**LT**	**Yes**	**LT processing**

TD: Time domain, FD: Frequency domain
ST: short-time, LT: long-time

Automatic SLR

- NIST LRE 2017 on 5 language clusters
- Training: previous LRE, Fisher & Switchboard corpora
- Testing: LRE 2017 Dev and Eval sets (NB (MLS14) and video speech (VS))
- SLR: UBM i-vector system

Table 2: SLR performance of different features: EER (%)

Features	DEV17		EVAL15
	MLS14	VS	MLS14
SDCC	10.22	6.49	11.82
IFCC	11.41	12.58	15.51
DBN	5.97	4.08	6.75
SDCC+IFCC	**7.15**	**5.32**	**9.44**
DBN+IFCC	**4.60**	**3.42**	**5.97**

Conclusions

- IFCC - LT information from analytic phase
- SDCC/DBN - LT information from spectral magnitude
- IFCC and SDCC/DBN are extracted using different speech processing strategies
- They contain complementary information
- Fusion of the complementary information benefits the SLR

References

1. A. Potamianos and P. Maragos, "Speech formant frequency and bandwidth tracking using multiband energy demodulation," The JASA, vol. 99, no. 6, pp. 3795-3806, 1996.
2. S.L. Marple Jr., "Computing the discrete-time "analytic" signal via FFT," IEEE Trans. Sig. Proc., vol. 47, no. 9, pp. 2600-2603, Sep 1999.
3. K. Vijayan, P. R. Reddy, and K. S. R. Murty, "Significance of analytic phase of speech signals in speaker verification," Speech Communication, vol. 81, pp.54-71, Jul 2016.

Presented at
2018 IEEE International Conference on Acoustics, Speech and Signal Processing, 15-20 April 2018, Calgary, Alberta, Canada.