Special Session: Phase Importance in Speech Processing Applications

Pejman Mowlaee, Rahim Saeidi, Yannis Stylianou

Signal Processing and Speech Communication (SPSC) Lab, Graz University of Technology
Speech and Image Processing Unit, School of Computing, University of Eastern Finland, Finland
Computer Science Dept. University of Crete, Crete, Greece

September 17, 2014

Special Session Programme at a Glance

1. Introductory talk (Special session aim and scope)
2. Shot-gun presentations
3. Poster presentation (poster boards: between 201 and 206)

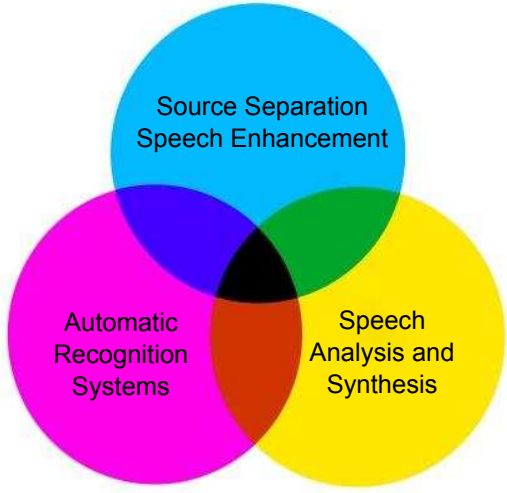
Time	Paper no.	Authors	Paper title
Oral 15 (min)	975	Pejman Mowlaee, Rahim Saeidi, Yanins Stylianou	INTERSPEECH 2014 Special Session on Phase Importance in Speech Processing Applications
Shot-gun 8 x 5 = 40 (min)	187	Estefania Cano, Mark Plumbley and Christian Dittmar	Phase-based harmonic/percussive separation
	218	Gilles Degottex and Nicolas Obin	Phase Distortion Statistics as a Representation of the Glottal Source: Application to the Classification of Voice Qualities
	219	Gilles Degottex and Daniel Erro	A measure of phase randomness for the harmonic model in speech synthesis
	707	Emma Jokinen, Marko Takanen, Hannu Pulakka and Paavo Alku	Enhancement of speech intelligibility in near-end noise conditions with phase modification
	719	S Aswin Shanmugam and Hema Murthy	A Hybrid Approach to Segmentation of Speech Using Group Delay Processing and HMM Based Embedded Reestimation
	738	Maria Koutsogiannaki, Olympia Simantiraki, Gilles Degottex and Yannis Stylianou	The Importance of Phase on Voice Quality Assessment
	896	Karthika Vijayan, Vinay Kumar and K. Sri Rama Murty	Feature Extraction from Analytic Phase of Speech Signals for Speaker Verification
	904	Jon Sanchez, Ibon Saratxaga, Inma Hernaez, Eva Navas and Daniel Erro	A Cross-vocoder Study of Speaker Independent Synthetic Speech Detection using Phase Information
65 min	Poster Presentation		

General Outline: Goal and Scope

- Demonstrating the importance of phase in different applications
- Consider the latest progress in phase-based speech processing
- Establish a new community of researchers working on phase

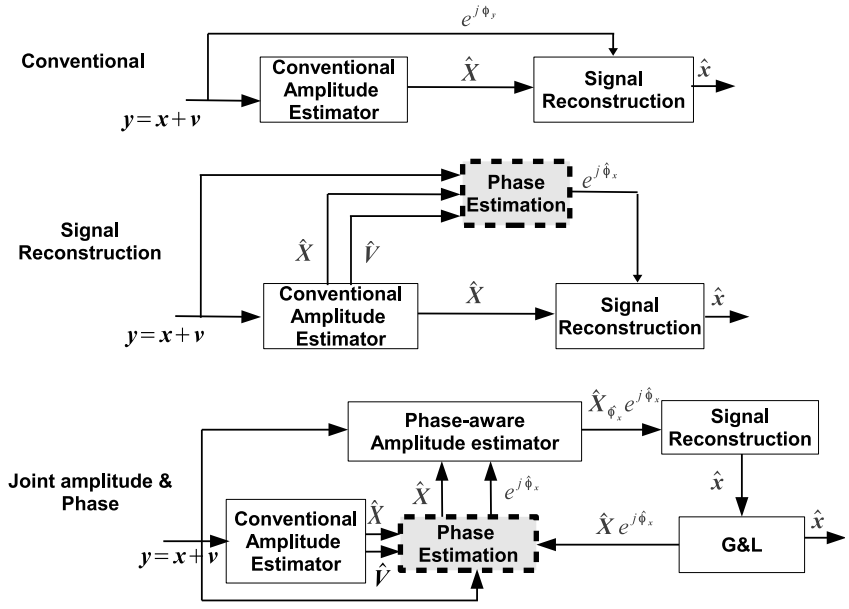
Overview on Phase Importance in Speech Applications

1. Source Separation and Speech Enhancement
2. Speech Analysis and Synthesis
3. Automatic Recognition Systems



Introduction to Speech Signal Enhancement & Research Questions

- Speech signals are impaired by additive noise, interfering speaker and reverberation
- Is phase important?[1,2] yes in both **amplitude estimation**[4,7] and **signal reconstruction**[3]. Impact on quality/intelligibility?[3—8]
- How to estimate clean phase? Is phase-aware signal enhancement possible?[4,7,8]



[1] D. Wang, J. Lim, "The unimportance of phase in speech enhancement," TASL 30(4), pp. 679-681, 1982.

[2] K. K. Paliwal et al., "The importance of phase in speech enhancement" Speech Communication, 53(4), pp. 465-494, 2011.

[3] P. Mowlaee, R. Saiedi, R. Martin, "Phase estimation for signal reconstruction in single-channel speech separation" INTERSPEECH 2012.

[4] T. Gerkmann and M. Krawczyk, "MMSE-optimal spectral amplitude estimation given the STFT-phase," IEEE SPL, 2013.

[5] P. Mowlaee, R. Martin, "On phase importance in parameter estimation for single-channel source separation," IWAENC 2012.

[6] C. Chacon, P. Mowlaee, "Least Squares Phase Estimation of Mixed Signals" INTERSPEECH, 2014.

[7] P. Mowlaee, R. Saeidi, "Iterative closed-loop phase-aware single-channel speech enhancement" IEEE SPL, 2013.

[8] K. Nathwani et al., "Group Delay Based Methods for Speaker Segregation and its Application in Multimedia Information Retrieval", TASL, 15(6) 2013.

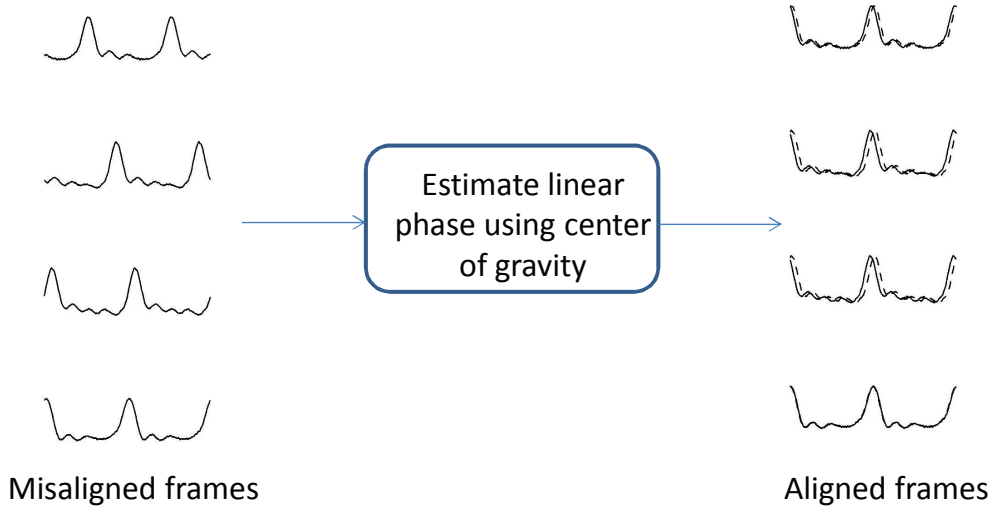
Phase/Amplitude Estimation From Noisy Speech

- Griffin & Lim based methods[2,3]
- Mowlaee et al. Geometry-based: using TF Constraints[4,5], Least Squares Phase Estimation[6]
- Gerkmann et al. phase-aware amplitude estimator[8]
- Mowlaee et al. Temporal smoothing of unwrapped phase[7]
- Sugiyama et al. Phase randomization[9]
- Mowlaee et al. Iterative closed-loop phase-aware[10]

[2] D. Griffin, J. Lim, "Signal Estimation from Modified short-time Fourier transform" TASL, 1984.
[3] P. Mowlaee et al., "Partial phase reconstruction using sinusoidal model in single-channel speech separation" INTERSPEECH, 2013.
[4] P. Mowlaee, R. Saeidi, R. Martin; "Phase Estimation for signal reconstruction in single-channel source separation" INTERSPEECH, 2012.
[5] P. Mowlaee, R. Saeidi; "Time-Frequency Constraints for Phase Estimation in Single-Channel Speech Enhancement" IWAENC, 2014.
[6] C. Chacon, P. Mowlaee, "Least Squares Phase Estimation of Mixed Signals" INTERSPEECH, 2014.
[7] J. Kulmer, P. Mowlaee, M. Watanabe, "A Probabilistic Approach for Phase Estimation in Single-Channel Speech Enhancement Using von Mises Phase Priors" MLSP, 2014.
[8] M. Krawczyk, T. Gerkmann, "STFT Phase Improvement for Single Channel Speech Enhancement" IWAENC, 2012.
[9] A. Sugiyama, R. Miyahara, "Phase randomization - a new paradigm for single-channel signal enhancement" ICASSP, 2013.
[10] P. Mowlaee, R. Saeidi, "Iterative closed-loop phase-aware single-channel speech enhancement" IEEE SPL, 2013.

Phase Importance in Speech Analysis/Synthesis

- Not used in an explicit way in Unit selection based TTS.
- Remove linear phase mismatches in concatenative speech synthesis using centre of gravity[1]
- Using minimum phase[2], leading to buzzness in synthesized quality esp. at fricatives.
- Introduce mixed phase in HMM-based TTS by suggesting complex cepstrum[2,3].
- Requires phase unwrapping: high dimension Fourier transform or linear phase removal.
- Possibility to include phase information without constraints in HMM?



[1] Y. Stylianou, "Removing linear phase mismatches in concatenative speech synthesis" TASL 9(3), 2001.
[2] R. Maia, M. Akamine, M.J.F. Gales, "Complex cepstrum as phase information in statistical parametric speech synthesis" ICASSP, 2012.
[3] R. Maia, Y. Stylianou, "Complex cepstrum factorization for statistical parametric synthesis" ICASSP, 2014

Phase-Based Processing for Recognition Applications

Challenges

- The starting point of a frame
- Phase wrapping

Phase information as for complementing amplitude information

- Feature-level fusion of amplitude and phase features
- Score-level fusion of recognition systems built on separate amplitude and phase features

Special application of phase information

- Detection of nasalized vowels from the mixture of oral and nasalized vowels
- Synthetic speech detection

Phase-Based Feature Extraction and Modeling

- Phase Evolution - R. Schlutel, and H. Ney
- Phase-Sensitive Model - L. Deng, J. Droppo, and A. Acero
- Phase AutoCorrelation - S. Ikbal, H. Hermansky, and H. Bourlard
- Residual Phase - K. Sri Rama Murty and B. Yegnanarayana
- (Modified) Group delay - R. M. Hegde, H. A. Murthy, and V. R. Gadde
- Relative Phase - L. Wang, S. Ohtsuka, S. Nakagawa
- Delta-Phase Spectrum (IF) - I. McCowan, D. Dean, M. McLaren, R. Vogt, S. Sridharan
- Teager Energy Operator (TEO) Phase - H. A. Patil, and K. Parhi
- ...

Conclusion

- Phase-based processing is a challenging topic still largely unsolved!
- 19 submissions (9 accepted)

Future outlook

- Special Issue in Speech Communication

Any Questions/Feedback?

Pejman Mowlaee (Email: pejman.mowlaee@tugraz.at)
Signal Processing and Speech Communication (SPSC) Lab, Graz, Austria
Rahim Saeidi (Email: rahim.saeidi@uef.fi)
Speech and Image Processing Unit, School of Computing, University of Eastern Finland, Finland
Yannis Stylianou (Email: yannis@csd.uoc.gr)
Computer Science Dept. University of Crete, Crete, Greece

Show and Tell: Iterative Refinement of Amplitude and Phase in Single-channel Speech Enhancement

- By: Pejman Mowlaee, Mario Watanabe, Rahim Saeidi
- When: Today 13:30-15:30 at Show and Tell Session 2