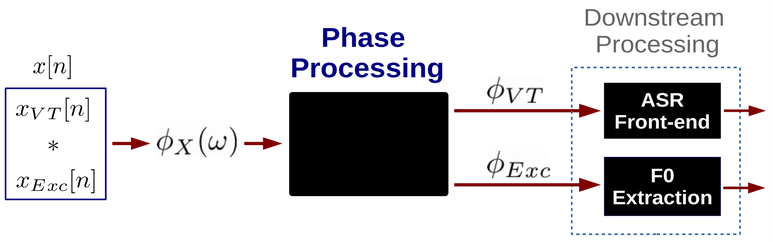
# Robust Source-Filter Separation of Speech Signal in the Phase Domain


**Erfan Loweimi, Jon Barker, Oscar Saz Torralba and Thomas Hain**

Speech and Hearing Research Group (SPandH), University of Sheffield, Sheffield, UK

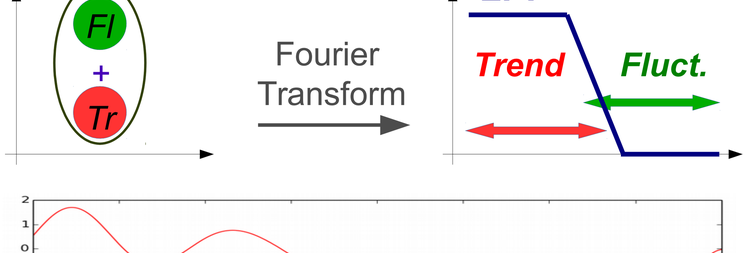
{e.loweimi1, j.p.barker, o.saztorralba, t.hain}@sheffield.ac.uk


## GOAL




## Signal Information Regions

$$
\begin{aligned}
X(\omega) &= X_{MinPh}(\omega)\, X_{AllP}(\omega) \\
|X(\omega)| &= |X_{MinPh}(\omega)| \\
arg[X(\omega)] &= arg[X_{MinPh}(\omega)] + arg[X_{AllP}(\omega)]
\end{aligned}
$$

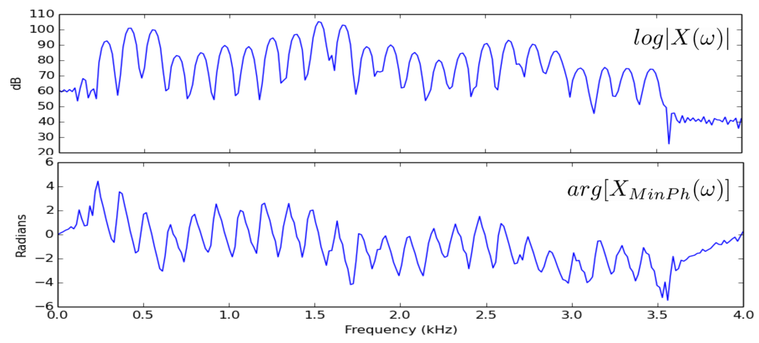



$$|X_{MinPh}(\omega)| \quad \underset{}{\overset{\text{HiL.Tran}}{\rightleftharpoons}} \quad \phi_{MinPh}(\omega)$$

## Source-Filter Separation in the Phase Domain

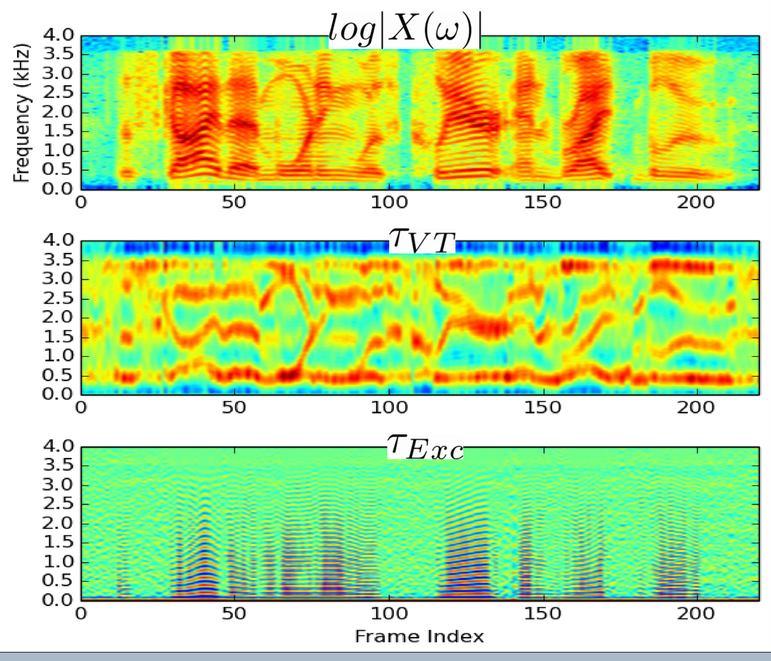
$$|X(\omega)| = |X_{VT}(\omega)|\,|X_{Exc}(\omega)| = |X_{MinPh}(\omega)|$$

$$
\begin{aligned}
arg[X_{MinPh}(\omega)] &= -\frac{1}{2\pi} log\big(|X_{VT}(\omega)|\,|X_{Exc}(\omega)|\big) * cot(\frac{\omega}{2}) \\
&= arg[X_{VT}(\omega)] + arg[X_{Exc}(\omega)]
\end{aligned}
$$

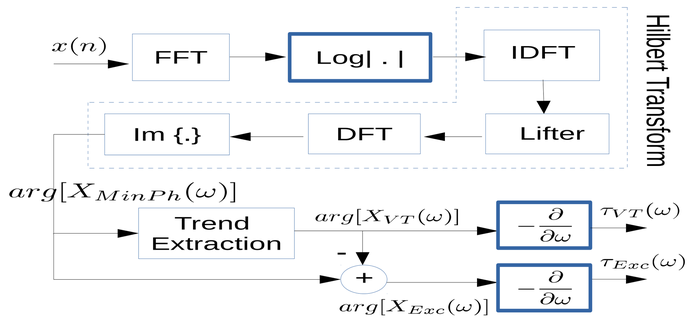


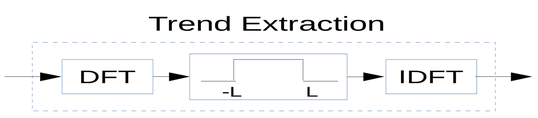

$$arg[X_{MinPh}] = \textit{Trend} + \textit{Fluctuation}$$




## The Proposed Workflow




## Improved Source-filter Separation in Phase Domain

## Generalised Logarithmic Function (GenLog)

$$
\begin{cases}
GenLog(x;\alpha) = \frac{1}{\alpha}(x^{\alpha} - 1), & x > 0 \quad \alpha \neq 0 \\
\lim_{\alpha \to 0} GenLog(x;\alpha) = log(x),
\end{cases}
$$

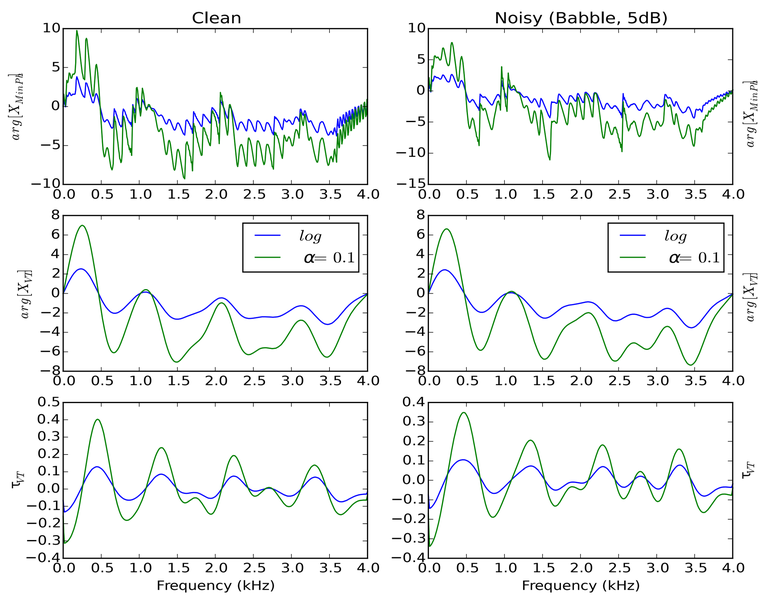
-- Statistics → Box-Cox Transformation (1964)

* Can improve the ...

*Linearity*, *Gaussianity* and *Homoscedasticity*

– Speech Processing → GenLog Function (1984)

* Can improve robustness to additive noise (Y?)
* Here → Dynamic range (DR) adjustment ...




$$arg[X_{MinPh}(\omega);\alpha] = -\frac{1}{2\pi\alpha}\underbrace{|X_{VT}(\omega)|^{\alpha}\,|X_{Exc}(\omega)|^{\alpha}}_{z^{\alpha}} * cot(\frac{\omega}{2})$$

$$
\begin{aligned}
z^{\alpha} = f(\alpha) &= 1 + \alpha\ logz + \alpha^2 (log\ z)^2 + ... \\
&\approx 1 + \alpha\ logz = 1 + \alpha\,(log|X_{VT}(\omega)| + log|X_{Exc}(\omega)|)
\end{aligned}
$$

$\alpha \gg 0 \rightarrow$ To supply a sufficient *DR*
$\alpha \ll 1 \rightarrow$ To avoid the violation of the quasi-additive combination of the *VT* and *Exc*

## Group Delay Properties

-- Disadvantage: Spikiness

++ Advantages:

– High spectral resolution / Low leakage
– Resembles the magnitude spectrum (Y?)

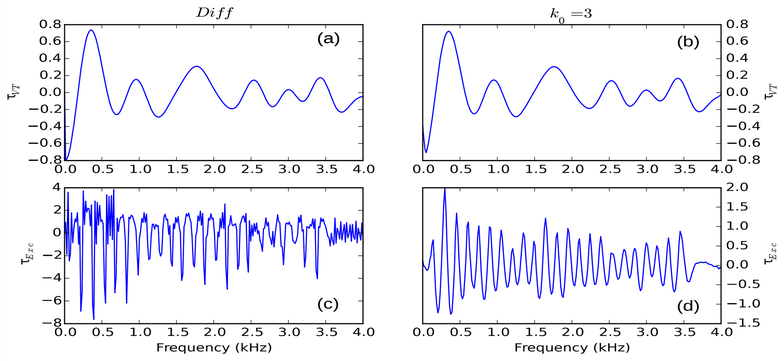
Differentiation acts as the *discriminator* in FM demodulator and *demodulates* information encoded in the phase spectrum

## Group Delay Computation

Numerical differentiation ...

– *Sample Difference* → intrinsically noisy
– *Regression filter* → context → Robustness

$$\tau_X[k] = -\frac{\sum_{m=-k_0}^{k_0} m\ \ arg[X[k+m]]}{\sum_{j=-k_0}^{k_0} j^2}$$

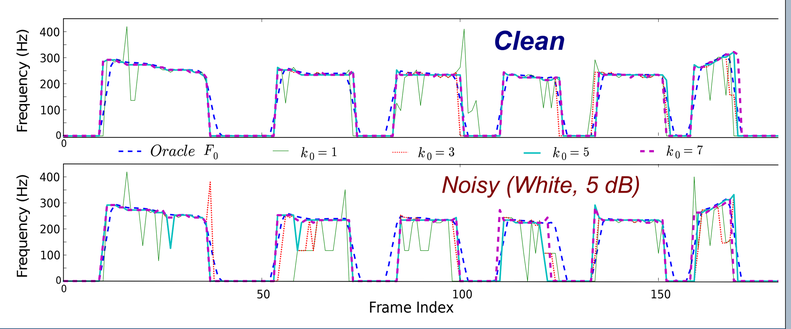



## Phase-based Pitch Extraction

Summation Harmonics Residuals (SRH)

$$SRH(k) = \tau_{Exc}(k) + \sum_{m=2}^{N_{harm}} [\tau_{Exc}(m\,k) - \tau_{Exc}((m-0.5)\,k)]$$

$$F_0 = \frac{f_s}{N_{FFT}} \underset{k}{argmax}\ SRH(k)$$

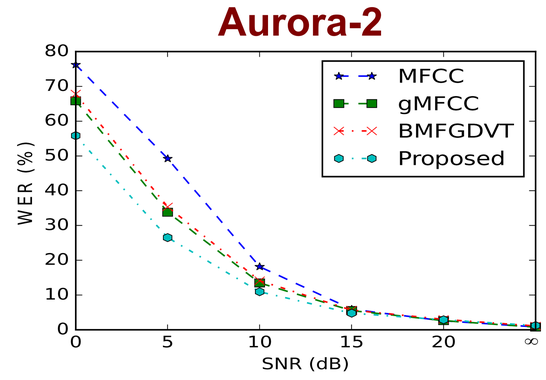



## ASR Results

| Feature | TestSet A | TestSet B | TestSet C |
|---|---|---|---|
| MFCC-E-D-A | 32.7 | 27.5 | 34.0 |
| $\gamma$-MFCC-0.075 | 24.6 | 23.8 | 23.1 |
| PLP | 32.7 | 29.4 | 33.8 |
| MODGD | 35.7 | 33.6 | 40.5 |
| ChirpGD | 33.0 | 27.0 | 40.6 |
| ProdSpec | 34.0 | 28.8 | 35.4 |
| ARGDMF | 24.6 | 21.0 | 24.0 |
| BMFGDVT (Baseline) | **26.8** | **22.6** | **26.6** |
| $\alpha$-BMFGDVT-0.12 | 22.8 | 20.8 | 20.7 |
| $\alpha$-BMFGDVT-0.1 | **22.1** | **19.5** | **20.5** |
| $\alpha$-BMFGDVT-0.08 | 22.3 | 19.3 | 21.0 |
| $\alpha$-BMFGDVT-0.1-1 | 21.7 | 19.2 | 20.3 |
| $\alpha$-BMFGDVT-0.1-2 | **21.5** | **18.9** | **20.3** |
| $\alpha$-BMFGDVT-0.1-3 | 21.5 | 18.8 | 20.5 |

Ave. 0-20 dB

**Aurora-2**




WER vs SNR

Ave. All Test Sets

**Aurora-4**

| Feature | A | B | C | D | *Ave* |
|---|---|---|---|---|---|
| MFCC [Clean] | 7.4 | 34.5 | 29.4 | 50.3 | 39.0 |
| Proposed [Clean] | 7.1 | 26.5 | 28.1 | 45.1 | 33.2 |
| BN-MFCC [Multi] | 7.2 | 12.9 | 14.3 | 27.0 | 18.7 |
| BN-Proposed [Multi] | 7.0 | 12.7 | 14.3 | 26.6 | 18.4 |