UMONS
Université de Mons

Phase-based Information for Voice Pathology Detection

T. Drugman, T. Dubuisson, T. Dutoit

POLYTECH MONS

fnrs
LA LIBERTÉ DE CHERCHER

Contents

- ✓ Introduction
- ✓ Phase-based features
- ✓ Experimental evaluation
- ✓ Conclusion

Contents

- ✓ **Introduction**
- ✓ Phase-based features
- ✓ Experimental evaluation
- ✓ Conclusion

Introduction

- **Most of speech processing techniques focus on the amplitude spectrum**

- **Phase-based features applied to speaker recognition and ASR** *Murty, 2006* *Hedge, 2004*

- **Promising approach for speech analysis**

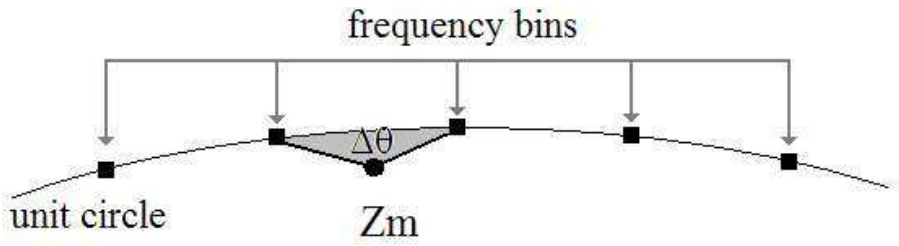
Contents

- ✓ Introduction
- ✓ **Phase-based features**
- ✓ Experimental evaluation
- ✓ Conclusion

Group delay-based analysis

Group delay is defined as:

$$\tau(\omega) = \frac{-d(\angle X(\omega))}{d\omega}$$



Group delay-based analysis

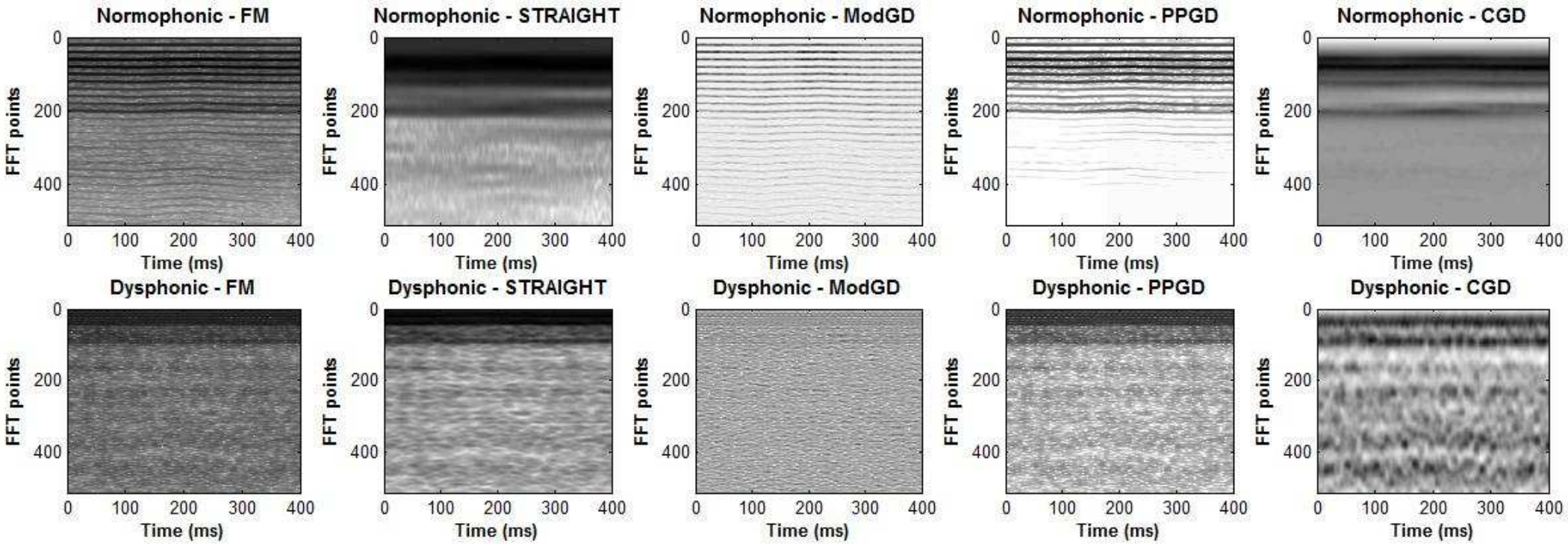
- **Modified Group Delay (MODGD)** *Hedge, 2004*

$$\tau_p(\omega) = \frac{X_R(\omega)\cdot Y_R(\omega) + X_I(\omega)\cdot Y_I(\omega)}{\left|S(\omega)\right|^{2\gamma}}$$

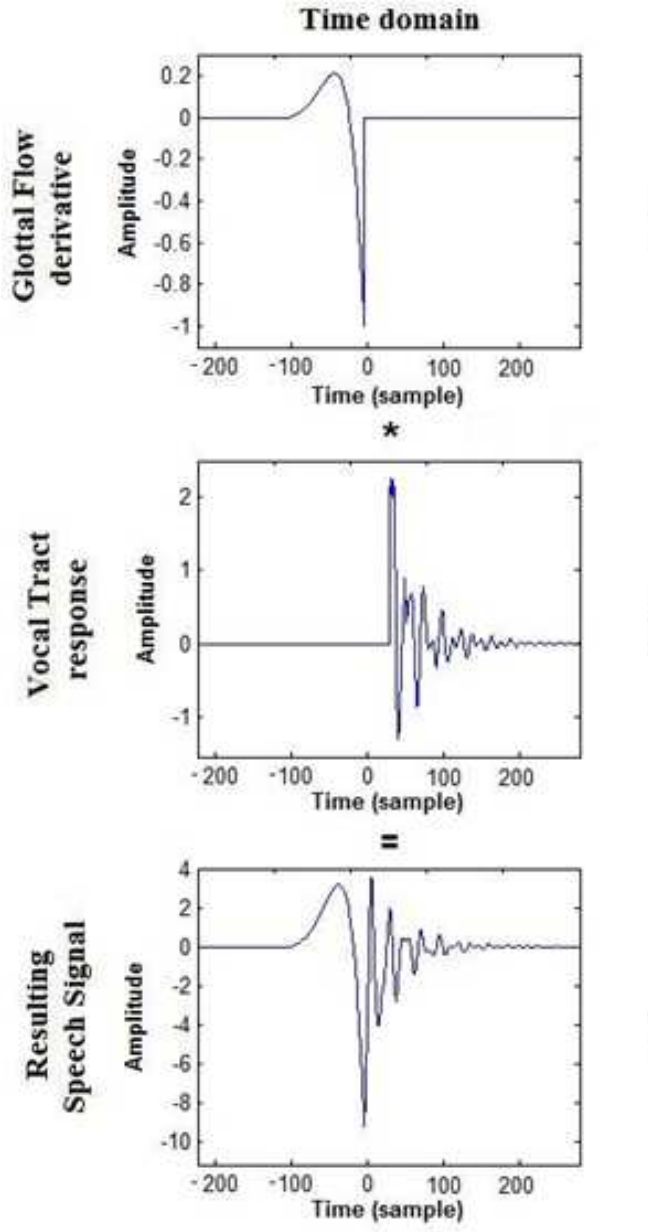
$$\tau_{\text{mod}}(\omega) = \frac{\tau_p(\omega)}{\left|\tau_p(\omega)\right|}\cdot\left|\tau_p(\omega)\right|^{\alpha}$$

Group delay-based analysis

- **Chirp Group Delay (CGD)** *Bozkurt, 2007*
- **STRAIGHT spectrogram** *Kawahara, 2002*
- **Fourier Magnitude (FM)**

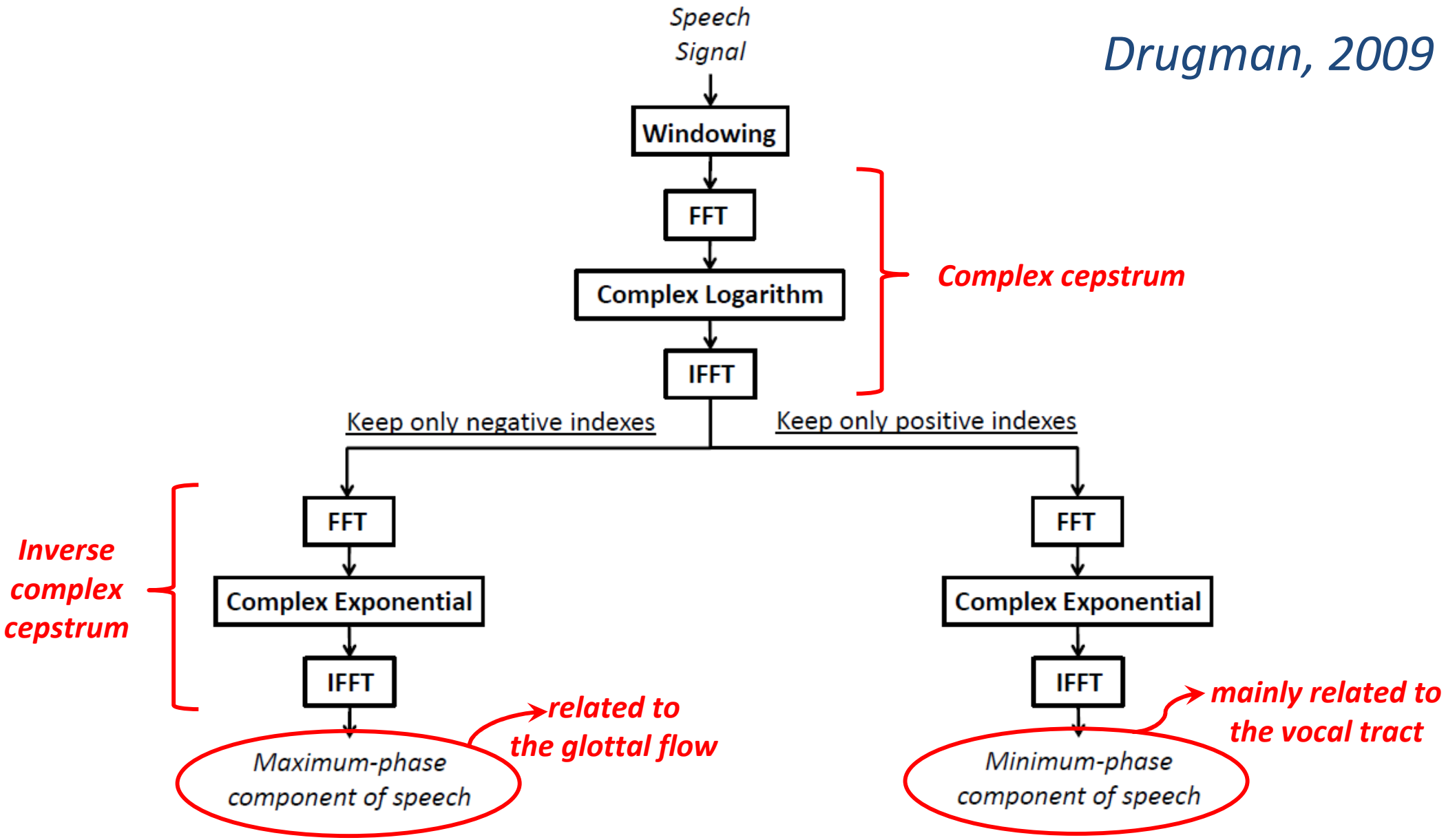


Mixed-Phase Decomposition

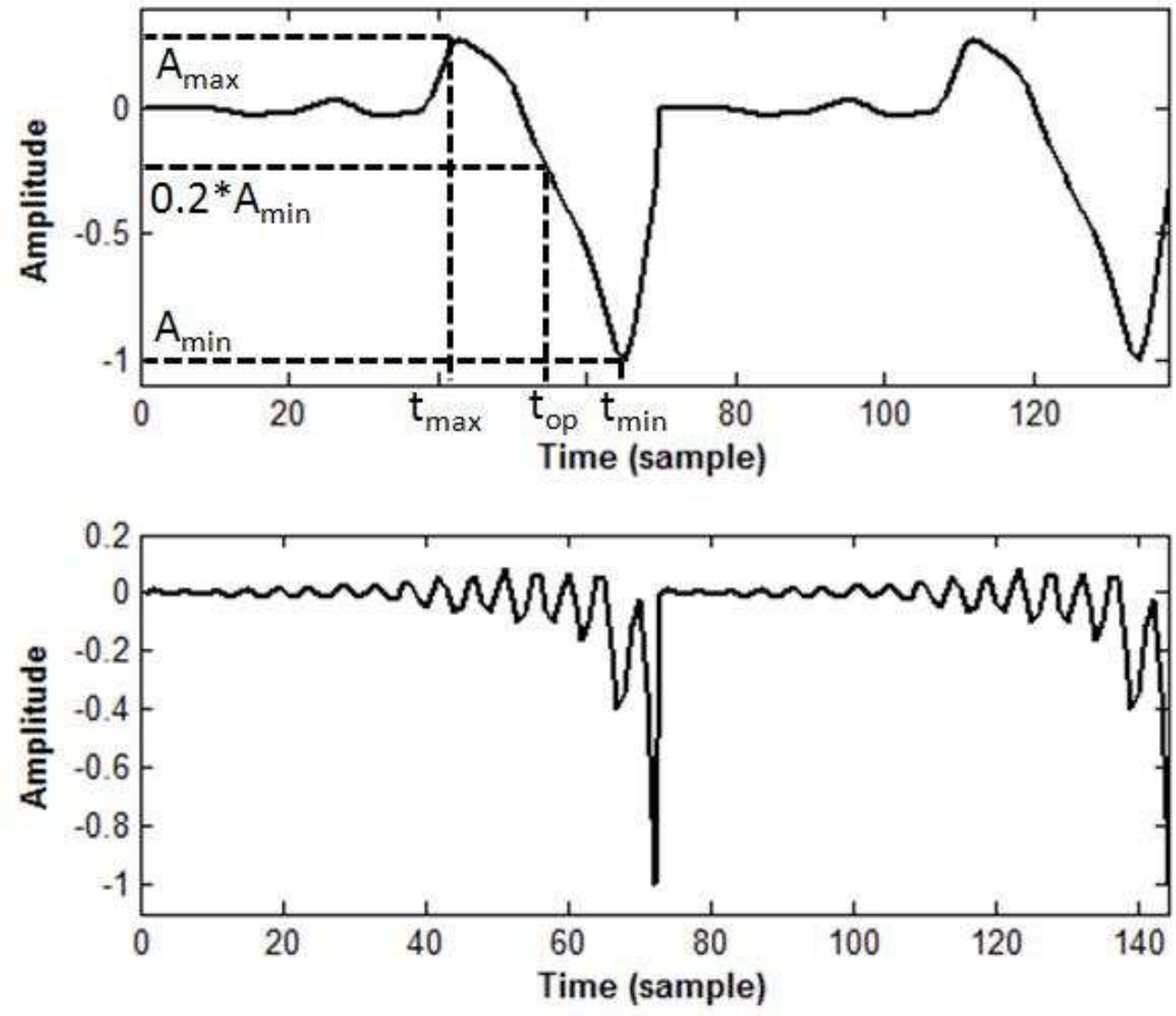


Complex Cepstrum Decomposition

Drugman, 2009



Adequacy of the mixed-phase model



$$T_1 = \frac{t_{\min} - t_{\max}}{T_0}$$

$$T_2 = \frac{t_{\min} - t_{op}}{T_0}$$

Contents

- ✓ Introduction
- ✓ Phase-based features
- ✓ **Experimental evaluation**
- ✓ Conclusion

Experimental Protocol

- **Database:** MEEI (Kay) database on sustained vowels (53 normo, 657 dyspho)

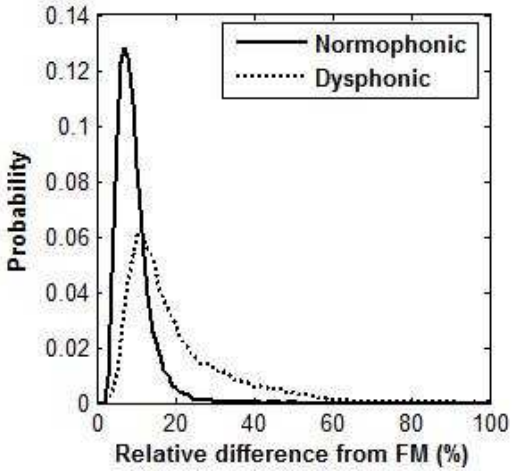
- **Features:**

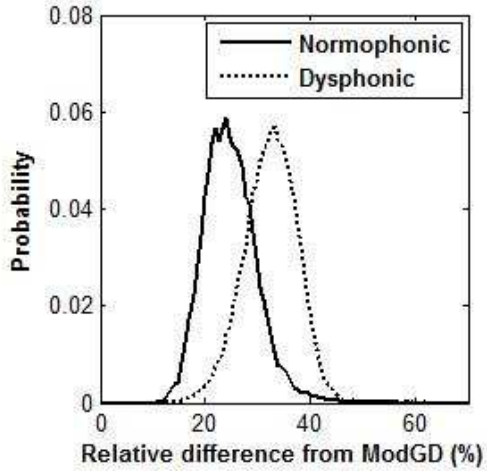
Frame-to-frame variations of the 5 spectrograms

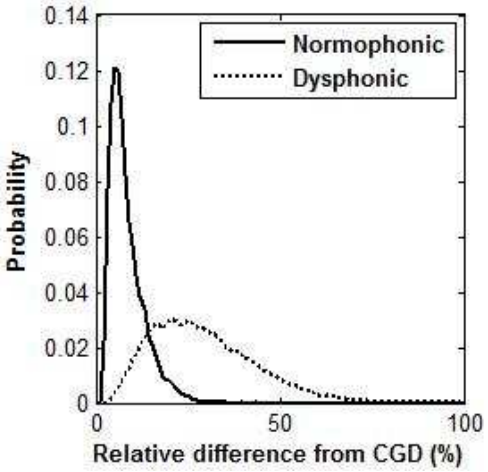
T1 and T2 for the mixed-phase model

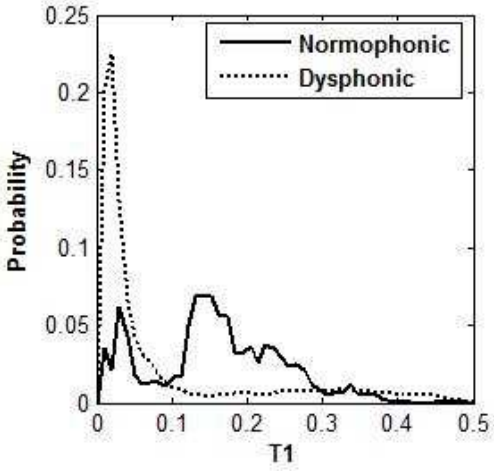
3 perceptual spectral balances from FM

Mutual Information-based Evaluation









Feature	dFM	dSTRAIGHT	dModGD	dPPGD	dCGD
Normalized MI (%)	22,32	16,32	30,56	15,43	55,97

Feature	T1	T2	Bal1	Bal2	Bal3
Normalized MI (%)	32,02	23,09	56,64	55,85	15,39

(Bal1,Bal2): 64.6% of MI **(Bal1,T2): 79.3% of MI**

Classifier-based evaluation

- ANN with 16 neurons (1 hidden layer)
- 10-fold cross validation
- Error rate at the frame and patient levels

Classifier-based evaluation

Features used	Error rate (frame level)	Error rate (patient level)
dFM	17.2	8.73
dCGD	9.40	4.93

Contents

- ✓ Introduction
- ✓ Phase-based features
- ✓ Experimental evaluation
- ✓ **Conclusion**

Conclusion

✓ **Phase-based features are suited for characterizing irregularities of phonation**

✓ **Phase-based features are complementary to parameters derived from the magnitude spectrum**

✓ **Good performance with only 3 features**