ICCT 2013

Paper ID: 8

Emotion Recognition from Speech using Teager based DSCC features

Santosh Chapaneri & Deepak Jayaswal
St. Francis Institute of Technology
University of Mumbai

Outline

- Motivation
- Emotion Recognition Overview
- Proposed Feature Extraction Technique
- Proposed Feature Recognition Technique
- Conclusion
- References

Why Emotion Recognition?

- Detecting frustration of callers to automated help line
- Computer tutorials via virtual avatars
- Lie detection
- Humanoid Robots

Basic Emotions

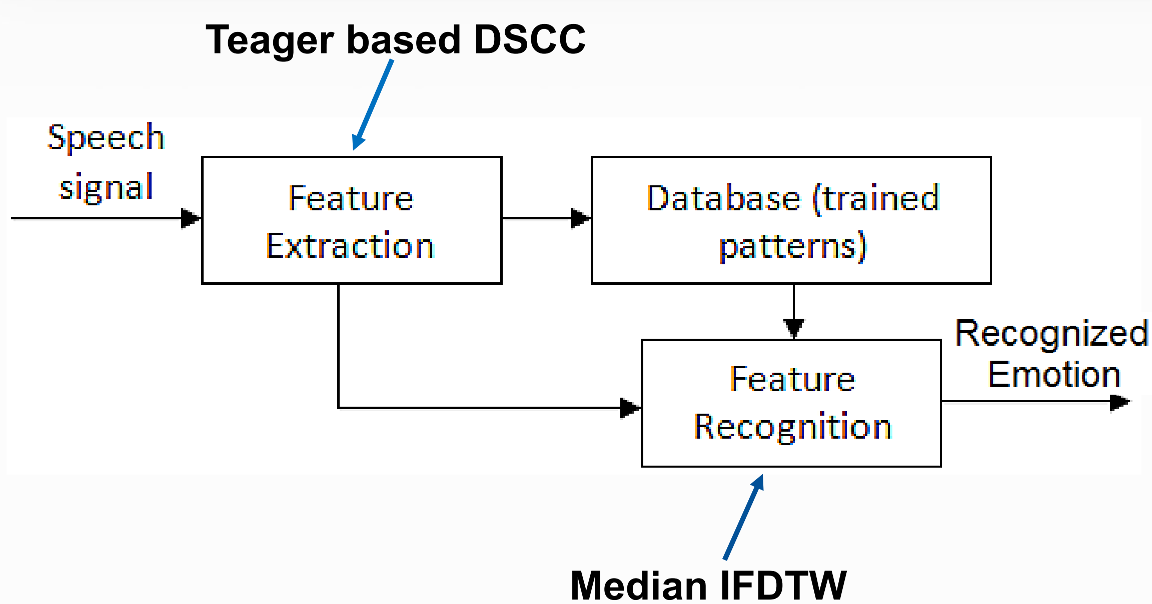
Speech Database

- Berlin Emotional Database (EMO-DB) [1]
- Total 535 utterances:

Anger	Boredom	Disgust	Fear	Happiness	Sadness	Neutral
127	81	46	69	71	62	79

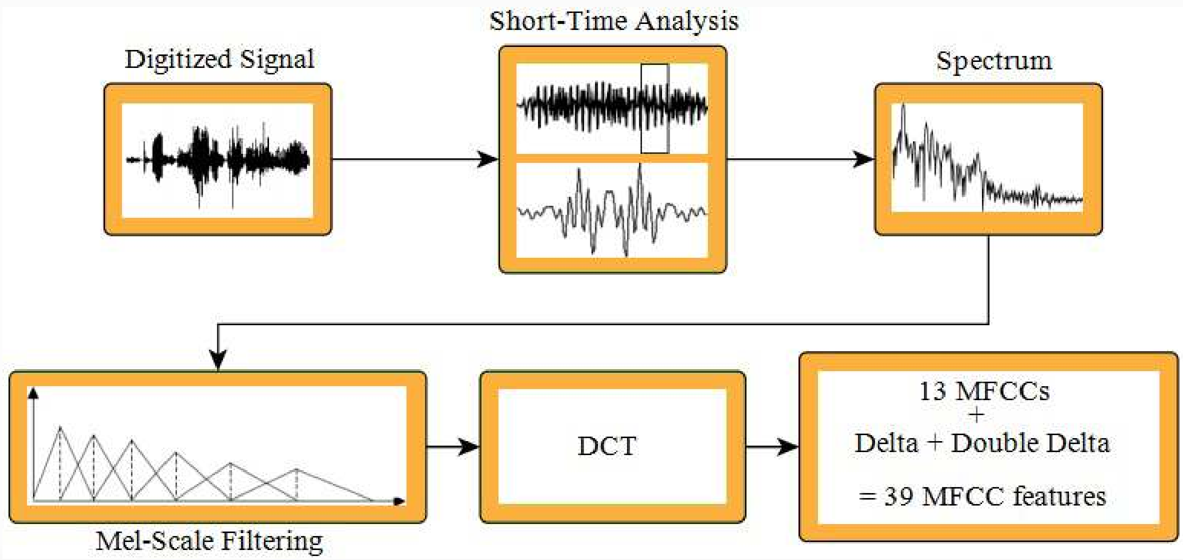
- 70% used for training, 30% for testing
- Sampling frequency 16KHz
- 16-bit resolution, mono channel samples

Emotion Recognition Overview



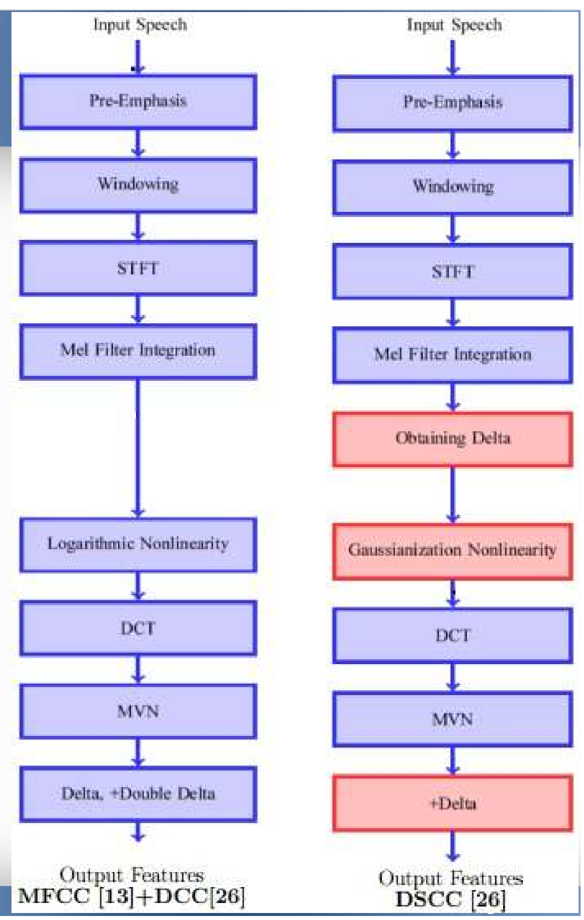
Conventional MFCC

Computation of Mel Frequency Cepstral Coefficients [2]:



DSCC Features

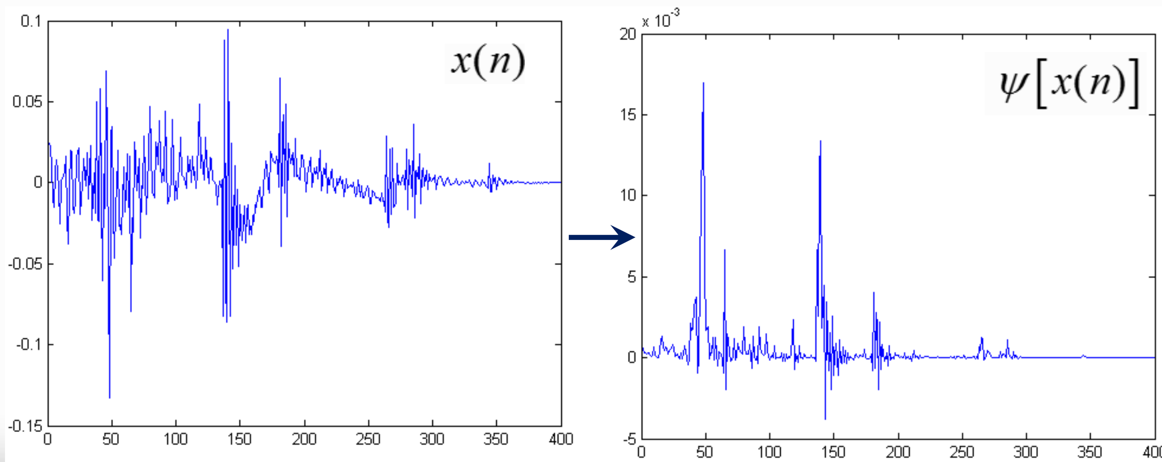
- Delta Spectral Cepstral Coefficients [3]
  - Delta in spectral domain instead of cepstral domain
  - Gaussian non-linearity instead of Log non-linearity
  - 2nd Delta over DSCC instead of MFCC features



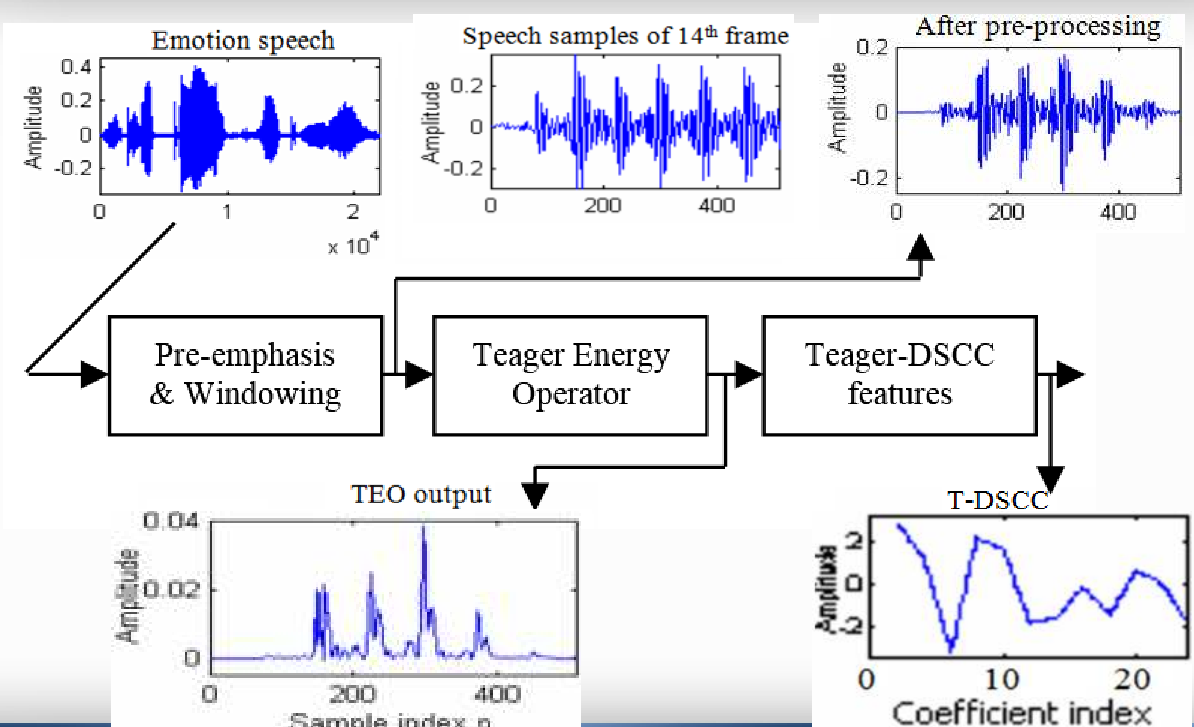
Teager Energy Operator (TEO)

- Non-linear energy tracking operator
- Useful mathematical model of vocal tract

$$\psi\left[x(n)\right] = x^2(n) - x(n+1)x(n-1)$$

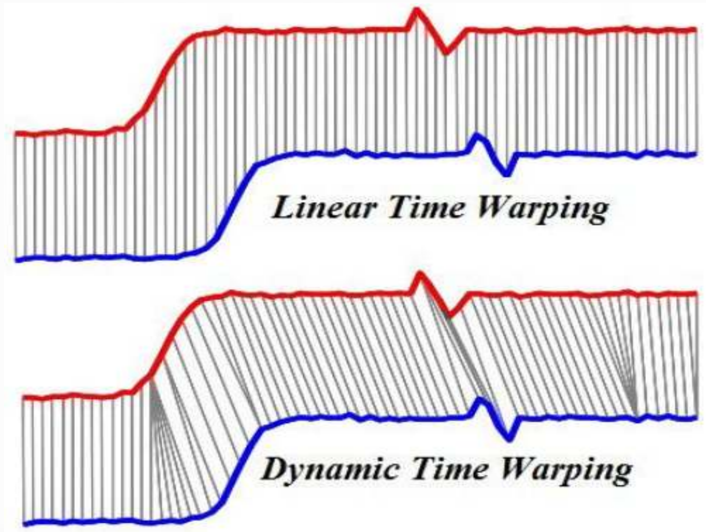


Proposed Feature Extraction



Feature Recognition

- Dynamic Time Warping (DTW)
- Warping - expanding/contracting the time dimension



DTW non-linear alignment of two time series

DTW Algorithm

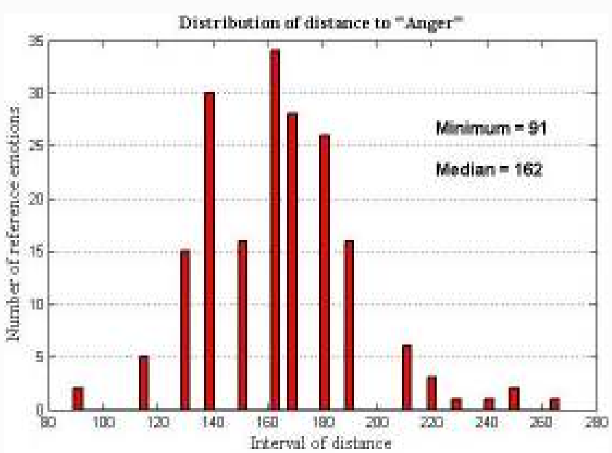
- Popular algorithm for automatic speech recognition based on template matching
- Dynamic programming approach based on Bellman’s optimality principle
- Solutions to slightly smaller problems used to find larger solutions

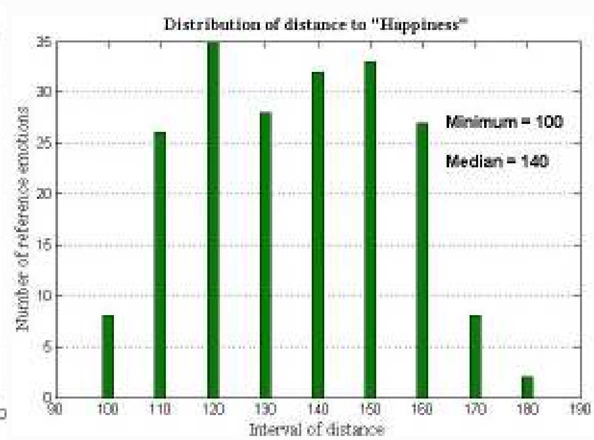
DTW Variants

- DTW ignores temporal relationship between feature values
- Derivative DTW (DDTW) [4]
- Improved Features for DTW (IFDTW) [5]
- Fast DTW [6] to reduce computations from O(N^2) to O(N)

Proposed Feature Recognition

- Distribution of distances between test emotion “Happiness” to reference emotions “Anger” & “Happiness”





- Median found to be accurate compared to Minimum distance
- Proposed Feature Recognition: Median Filtered IFDTW

Experimental Results

- Clean speech + Speech under 10 dB Gaussian noise
- 161 utterances for testing against 374 trained samples

Overall Recognition Accuracy (%)

	#Features	DTW	IFDTW	Median IFDTW
MFCC + $\Delta + \Delta\Delta$	39	84.52	87.39	91.29
DSCC	26	93.82	95.14	96.73
T-DSCC	**26**	94.18	95.69	**97.52**

Experimental Results

Confusion Matrix (23 utterances/emotion)
(A: Anger, B: Boredom, D: Disgust, F: Fear, H: Happiness, S: Sadness, N: Neutral)

	A	B	D	F	H	S	N
A	**23**	0	0	0	0	0	0
B	0	**23**	0	0	0	0	0
D	0	0	**22**	0	0	0	1
F	0	0	0	**23**	0	0	0
H	0	0	0	0	**23**	0	0
S	0	1	0	1	0	**21**	0
N	0	1	0	0	0	0	**22**

Conclusion

- Proposed Teager-DSCC features are robust to noise in input speech signal
- Proposed Median Filtered IFDTW gives accurate classification of test emotion
- Overall recognition accuracy higher compared to existing systems

- **Future scope**:
  - Indian native speech
  - App for Aakash tablets

Thank You

References

[1]	F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, “A database of German emotional speech”, Proc. Interspeech-2005, Lisbon, Portugal, pp. 1-4, Jan 2005
[2]	L. Rabiner, and B. Juang, Fundamentals of Speech Recognition, Prentice Hall, 1993
[3]	K. Kumar, C. Kim, and R. Stern, “Delta-Spectral Cepstral Coefficients for robust speech recognition”, IEEE Intl. Conf. Acoustics, Speech, and Signal Processing, pp. 4784-4787, May 2011
[4]	E. Keogh, and M. Pazzani, “Derivative dynamic time warping”, Proc. of the 1st SIAM Intl. Conf. Data Mining, Chicago, USA, 2001
[5]	S. Chapaneri, “Spoken digits recognition using weighted MFCC and improved features for dynamic time warping”, Intl. Journal Computer Applications, vol. 40, no. 3, pp. 6-12, Feb 2012
[6]	S. Salvador, and P. Chan, “FastDTW: toward accurate dynamic time warping in linear time and space”, Proc. 3rd KDD Workshop on Mining Temporal and Sequential Data, pp. 70-80, Aug 2004
[7]	http://www.takanishi.mech.waseda.ac.jp/top/research/we/we-4rii/index.htm