BIOLOGICALLY INSPIRED SPEECH EMOTION RECOGNITION

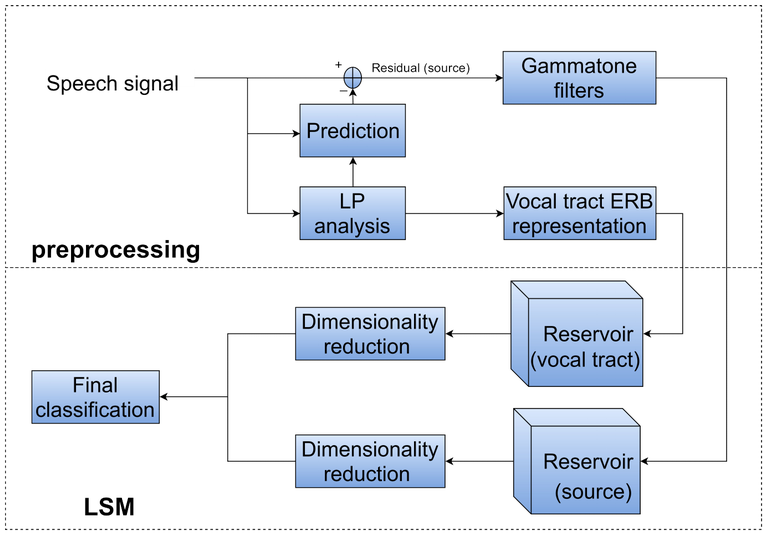
Reza Lotfidereshgi, Philippe Gournay
{Reza.Lotfi.Dereshgi, Philippe.Gournay}@USherbrooke.ca

Université de Sherbrooke
Speech and Audio Research Group
2500, boul. de l'Université
Sherbrooke (Québec)
J1K 2R1 Canada

1. Summary

- Emotions are critical in speech communication
- The precise set of spectral and prosodic features for their automatic recognition is unknown
- A biologically-inspired system that operates on raw speech data and thus requires no feature extraction is proposed
- State of the art performance is achieved

2. Overview of the system

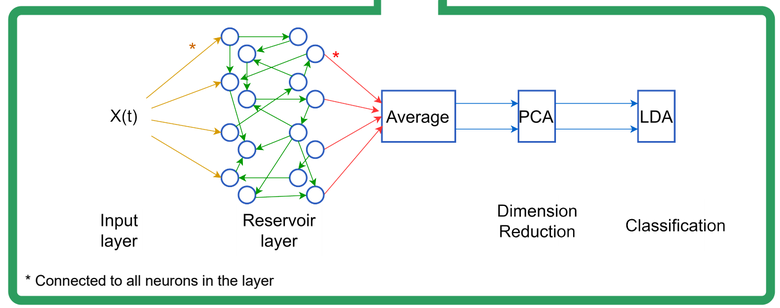


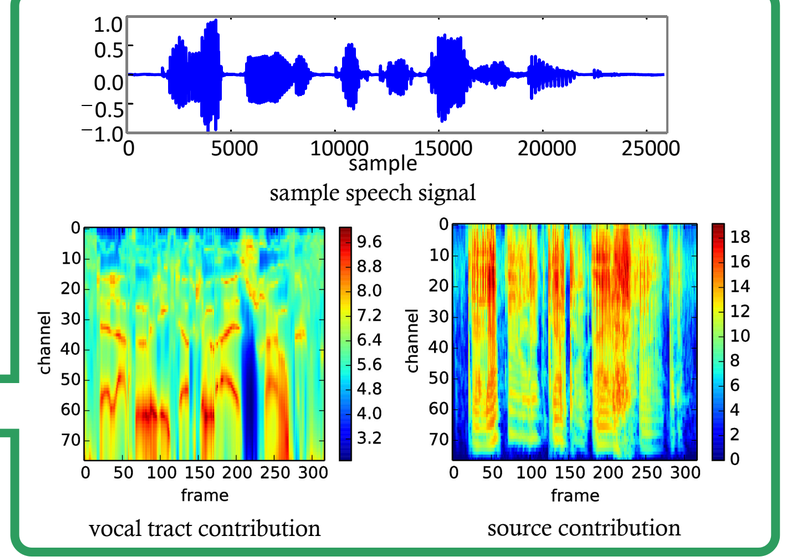
3. Preprocessing

- **LP analysis:** An order 16 LP filter is computed every 5ms using a 30ms Hamming window
- **Source contribution:** Log-energy of 5ms segments of each output of a 77-channel gammatone filterbank
- **Vocal tract contribution:** Frequency response of each LP filter, on an Equivalent Rectangular Bandwith (ERB) scale, in dB

4. Liquid State Machine design

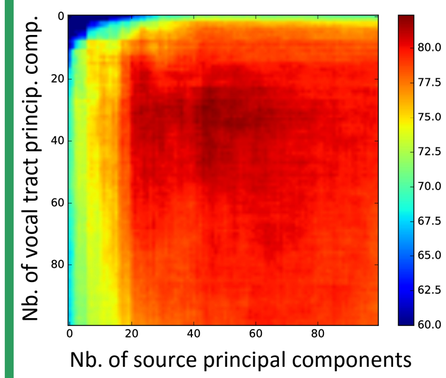
- 77*3*3 neurons arranged in a 3D structure with local synapse connections to create a tonotopic structure
- Integrate-and-fire neurons inspired from biological cortical neurons
- Asymmetric Spike Time-Dependent Plasticity (STDP) models of synaptic plasticity tuned for vocal tract and source reservoirs separately
- Dimensionality reduction by Principal Component Analysis (PCA) of the average activity of the neurons
- Final stage is Linear Discriminant Analysis (LDA)

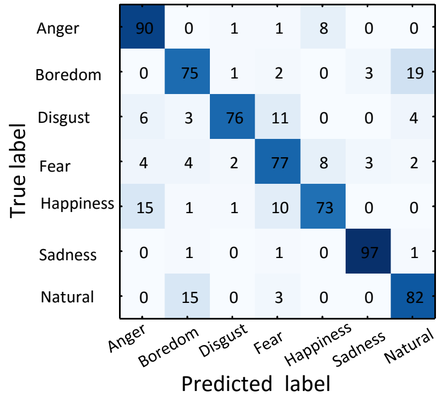




5. Performance Evaluation

- The proposed method was tested on the Berlin Database of Emotional Speech (Emo-DB)
- The highest recognition rate of 82.35% was achieved for 29 vocal tract and 44 source principal components
- This rate falls to 75.73% for a single-reservoir classifier with no source and vocal tract separation





6. Conclusions

- The proposed method operates directly on the speech signal, thus requires no feature extraction
- It is also largely inspired by biology:
  - Separating source and vocal tract contributions builds upon the motor theory of human speech perception
  - Both these contributions are analyzed on the perceptually-relevant ERB scale
  - Spiking neurons and synaptic plasticity models are very close to actual cortical neurons
- Outperforms current methods on comparable data

ICASSP 2017, New Orleans, Louisiana, USA