TOWARD COMPUTATIONAL MODELS OF SOUND PROCESSING

Stéphane Loiselle — Ramin Pichevar — Jean Rouat

Groupe de Recherche en Neurosciences Computationnelles et Traitement Intelligent des Signaux
Département de génie électrique et génie informatique, Université de Sherbrooke

Stephane.Loiselle@usherbrooke.ca — Ramin.Pichevar@usherbrooke.ca — Jean.Rouat@usherbrooke.ca

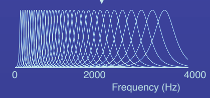
ABSTRACT

From physiology we learn that the auditory system extracts simultaneous features from the underlying signal, giving birth to simultaneous representations of audible signals. We also learn that pattern analysis and recognition are not separated processes (in opposition to the engineering approach of pattern recognition where analysis and recognition are usually separated processes).

SYNCHRONIZATION: SOURCE SEPARATION

Signal mixture

Gammatone filterbank

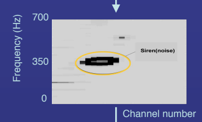


Envelope extraction

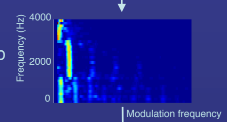
FFT

FFT

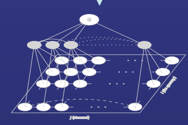
Cochleotopic spectrotopic map (CSM)



Amplitude modulation map (CAM)



Channel binding RN-Spikes

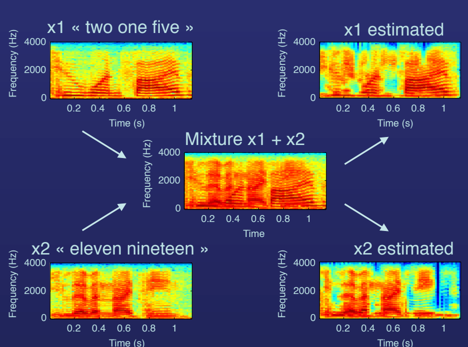


Mask generation and resynthesis

Reconstruction

In a first application, we combine a simultaneous auditory image representation with a network of oscillatory spiking neurons to segregate and bind auditory objects for acoustical source separation. It is shown that the spiking neural network performs unsupervised auditory image segmentation (to find 'auditory' objects) and binding of the objects belonging to the same auditory source (yielding automatic sound source separation).

EXAMPLE OF SEPARATION

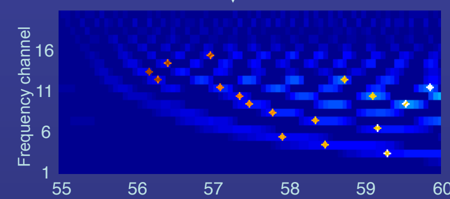


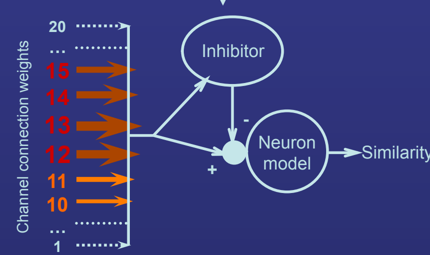
RANK ORDER CODING: SPEECH RECOGNITION

Rectification

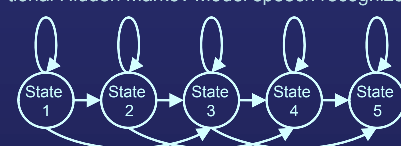
Compression

Thresholding





In the visual system, it has been observed that the order in which a population of cells fire is crucial to perform fast visual recognition tasks (Rank Order Coding). The use of the ROC has also been recently hypothesized in the mammalian auditory system. In this application we compare a very simplistic speech recognition prototype that uses the ROC with a conventional Hidden Markov Model speech recognizer.



Hidden Markov models.

EXPERIMENTS AND RESULTS

Speech database

- In house speech database made of 10 French digits (from 0 to 9) spoken ten times by 5 men and 4 women

Training and recognition

- For each digit, two reference models are used for the recognizer (one pronunciation for each sex).

Recognition results

- Conventional HMM system: **51.9%**
- Prototype: **64.89%**

RECOGNITION FOR EACH PRONUNCIATION OF THE TEN FRENCH DIGITS (PROTOTYPE).

#	Models										%
	1	2	3	4	5	6	7	8	9	0	**64.89**
1	84	1	4							1	**93.33**
2		69	2	1				3	13	2	**76.67**
3	10		58	18				1	1	2	**64.44**
4			22	68							**75.56**
5	11		1	2	42	21	6			7	**46.67**
6					1	68	13	2	1	5	**75.56**
7		1	1	2	11	49	12	1	1	12	**13.33**
8		1						68	16	5	**75.56**
9			4					9	54	23	**60**
0						15	2	3	9	61	**67.78**

ACKNOWLEDGMENT

Many thanks to Emmanuel Tessier and H. H. Anh.

REFERENCES

- S. Loiselle, J. Rouat, D. Pressnitzer and S. Thorpe, "Exploration of Rank Order Coding with Spiking Neural Networks for Speech Recognition," in Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, July 31 - August 4, 2005, pp. 2076-2080.
- S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system," Nature, vol. 381, pp. 520–533, june 1996.
- R. VanRullen, R. Guyonneau, and S. J. Thorpe, "Spike times make sense," Trends in Neurosciences, vol. 28, no. 1, p. 4, january 2005.
- J. Rouat, R. Pichevar, "Source separation with one ear: Proposition for an anthropomorphic approach," EURASIP Journal on Applied Signal Processing (Special Issue on Anthropomorphic Processing of Audio and Speech, invited), no. 9, pp 1365-1374, 2005