HAL
open science

Auditory bubbles reveal sparse time-frequency cues subserving identification of musical voices and instruments

Vincent Isnard, Clara Suied, Guillaume Lemaitre

To cite this version:

Vincent Isnard, Clara Suied, Guillaume Lemaitre. Auditory bubbles reveal sparse time-frequency cues subserving identification of musical voices and instruments. Meeting of the Acoustical Society of America, 2016, Honolulu, United States. pp.3267, 10.1121/1.4970361 . hal-01466181

HAL Id: hal-01466181

https://hal.science/hal-01466181v1

Submitted on 13 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Auditory bubbles reveal sparse time-frequency cues subserving identification of musical voices and instruments

Vincent Isnard [1,2] (vincent.isnard@ircam.fr), Clara Suied [2], Guillaume Lemaitre [1]

[1] *Institut de Recherche et Coordination Acoustique/Musique (IRCAM), CNRS UMR 9912, Equipe Perception et Design Sonore, Paris*

[2] *Institut de Recherche Biomédicale des Armées (IRBA), Département Action et Cognition en Situation Opérationnelle, Unité Perception, Brétigny-sur-Orge*

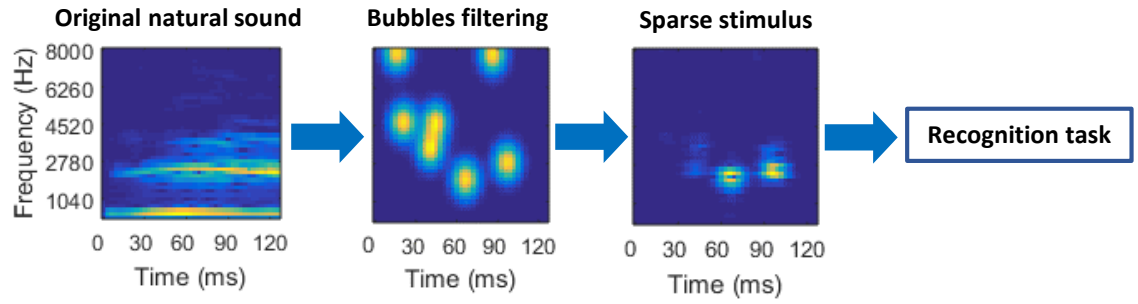
Introduction:

- Human listeners identify effortlessly natural sounds.
- What are the auditory cues underlying the recognition of natural sounds?

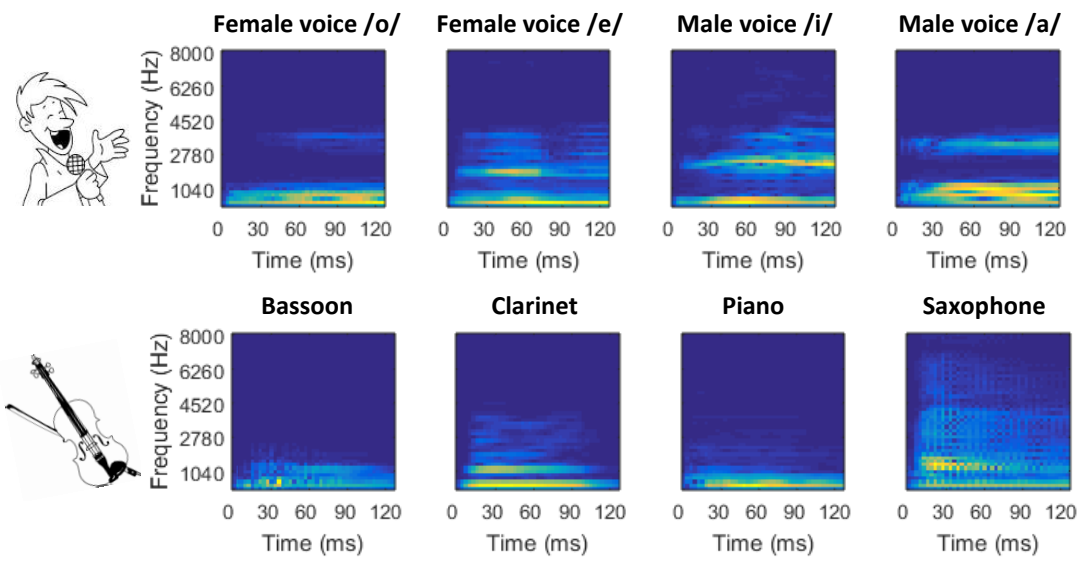
Identification of sparse time-frequency patterns subserving timbre recognition of natural sounds.

- Perceptual task vs. computational model.

Rationale: adapt a random search method called "Bubbles" proposed in vision (Gosselin & Schyns, 2001).

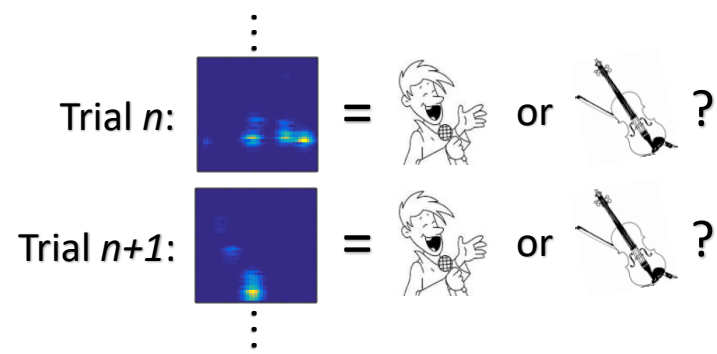


Original natural sounds: musical voices and instruments with same pitch (F#4) and same duration (128 ms with the attack).



Procedure: at each trial:

- Random selection of one sound → bubbles filtering.
- **2-AFC task:** Voice or Instrument?



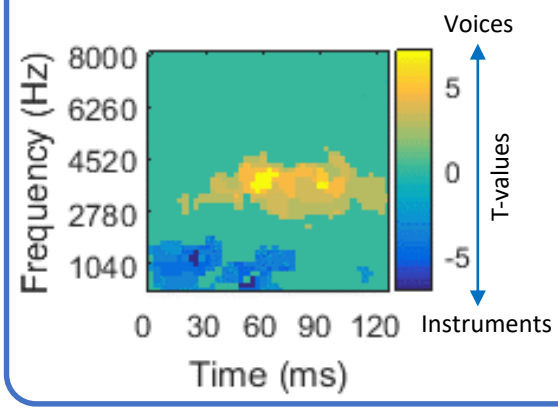
Decision:

- **Humans:** 8 participants; 1500 trials.
- **Computational model:** auditory distances between the sparse stimulus and the original sounds (cf. Isnard et al., 2016); 6400 trials.

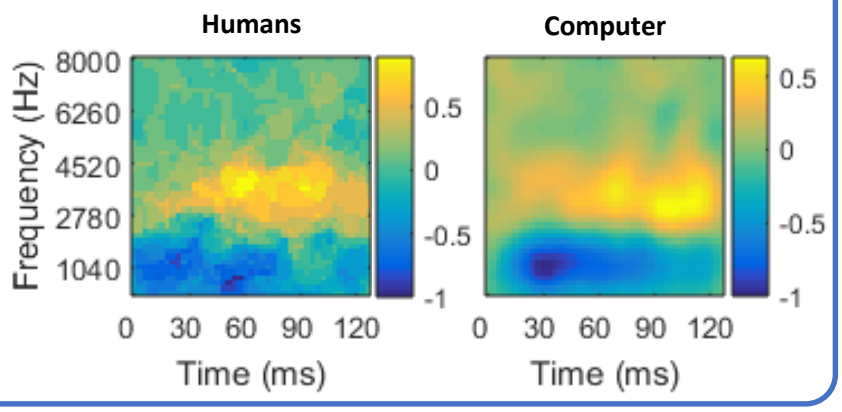
Results:

Humans	Computer
Sensitivity and bias	
▪ d' = 1.49 (± 0.40) ▪ c = 1.04 (± 0.38) (bias toward voices)	▪ d' = 1.47 ▪ c = 0.51
Auditory Classification Images (ACIs) computations	
▪ **ACI:** normalized mean correct image by participant and for each sound category, then subtraction. ▪ **Permutation test:** 1000 permutations by participant. ▪ **Thresholding** of each ACI's time-frequency bin compared to the 95th percentile of its permutation distribution. ▪ **T-tests** on each TF bin, then FDR for multiple hypothesis testing ($q < 0.05$).	▪ **ACI:** 400 correct responses by category with the highest auditory distances, then subtraction.

Human thresholded t-ACI



ACI comparison: humans vs. computer



Background:

- Bubbles adapted very recently in the auditory domain by few authors studying speech intelligibility (Mandel et al., 2016; Venezia et al., 2016).
- Using this technique, we follow the recent view of sparse auditory perception allowing the recognition of natural sounds (see Isnard et al., 2016).

Discussion:

- **Main result:** sparse features allow timbre recognition in particular:
  - Voice recognition = formant recognition.
  - Instrument recognition = attack recognition in lower frequencies.
- **Auditory distance model between natural sound categories:**
  - Striking similarity with the ACI obtained with human participants.
  - Auditory recognition of sparse stimuli seems to rely on the comparison with referent sounds from the different auditory categories implicated in the task.

References.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision research, 41*(17), 2261-2271.

Isnard, V., Taffou, M., Viaud-Delmon, I., & Suied, C. (2016). Auditory sketches: very sparse representations of sounds are still recognizable. *PloS one, 11*(3), e0150313.

Mandel, M. I., Yoho, S. E., Healy, E. W. (2016). Measuring time-frequency importance functions of speech with bubble noise. *The Journal of the Acoustical Society of America, 140*(4), 2542-2553.

Moore, B. C. (2003). Temporal integration and context effects in hearing. *Journal of Phonetics, 31*(3), 563-574.

Pressnitzer, D., Agus, T., & Suied, C. (2015). Acoustic timbre recognition. *Encyclopedia of Computational Neuroscience*, 128-133.

Venezia, J. H., Hickok, G., & Richards, V. M. (2016). Auditory "bubbles": efficient classification of the spectrotemporal modulations essential for speech intelligibility. *The Journal of the Acoustical Society of America, 140*(2), 1072-1088.