

Premiers résultats d'un projet sur la directivité de la voix

P. Luizard^{1,4}, H. Demontis², P. Stitt², F. Brinkmann¹, J.M. Arend¹, M. Schneider³,
B.F.G. Katz², F. Ollivier², S. Weinzierl¹, F. Silva⁵

¹Audio Communication Group, Technische Universität Berlin, Germany

²Sorbonne Université, CNRS, UMR 7190, Institut Jean Le Rond d'Alembert, Paris, France

³Georg Neumann GmbH, Berlin, Germany

⁴Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

⁵Aix Marseille Univ, CNRS, Centrale Marseille, LMA, Marseille, France

Introduction et motivation

Mesures de directivité conduites au 20^{ème} siècle [Trendelenburg, 1929 → Chu & Warnock, 2002]

Qualité et nombre croissant de microphones

21^{ème} siècle: systèmes de mesure plus denses, circulaires [Brandner et al, 2018] et sphériques [Arend et al, 2019]

Grand nombre de microphones pour mesures sur humains → MEMS microphones? [Demontis et al, 2018]

Grande différences de directivité à l'échelle des phonèmes, i.e. au sein des mots [Kocon, 2018]

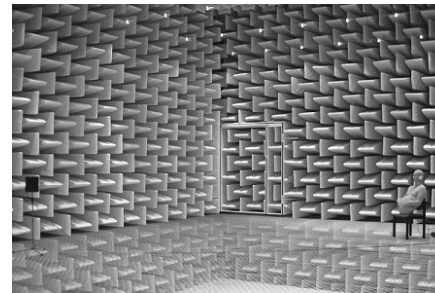
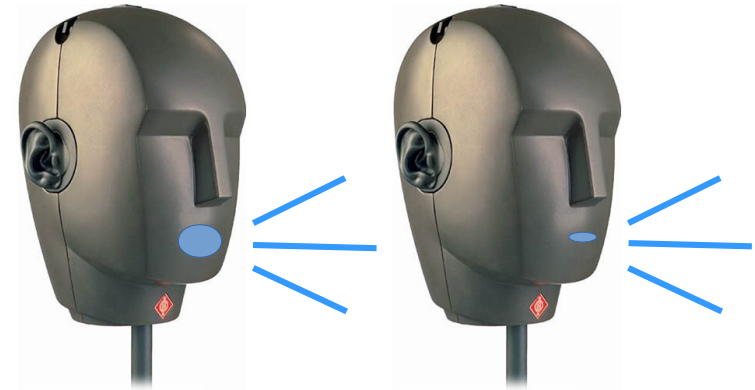
Hypothèse: Une **résolution spatiale élevée** devrait permettre d'observer les **détails de la directivité de la voix**

→ **Quelles informations peut-on capturer avec une antenne circulaire de MEMS? (largeur de spectre, dynamique)**

(Quel degré de résolution spatiale est nécessaire pour décrire complètement la directivité de la voix humaine?)

Organisation du projet

- Mesures sur des humains
(Institut d'Alembert, Paris
LMA, Marseille)
 - dans la bouche
 - champ proche (~ cm)
 - champ lointain (~ m)
- Construction d'une tête artificielle
(Georg Neumann GmbH, Berlin)
- Mesures sur cette tête artificielle
- Expériences perceptives
(TU Berlin
LMA, Marseille)



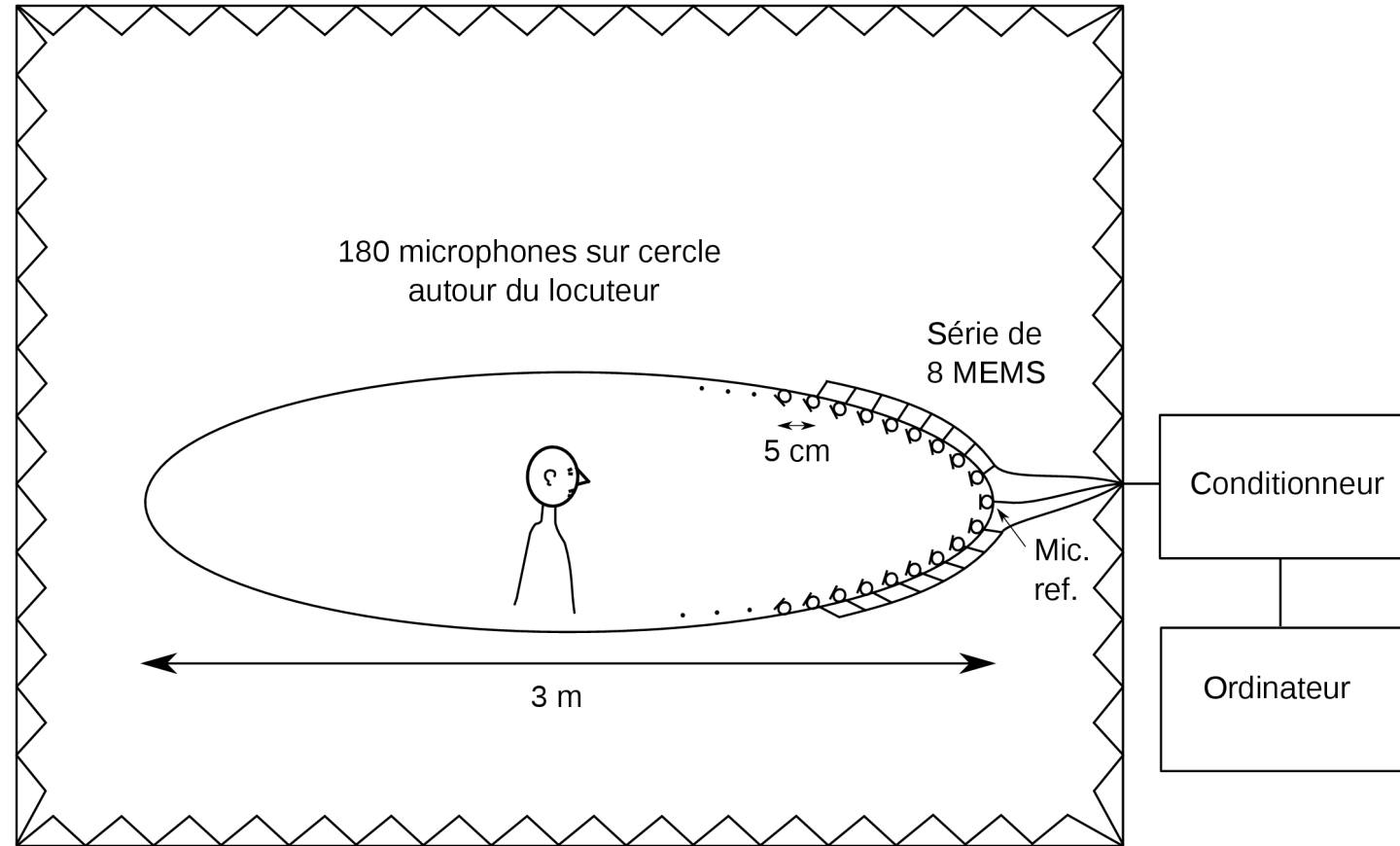
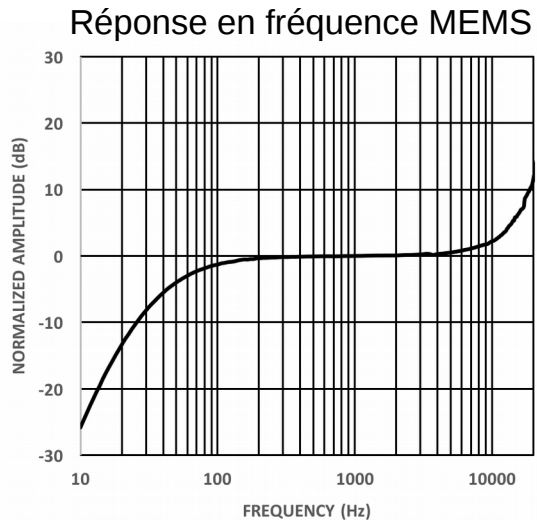
Mesures - setup

Chambre anéchoïque,
Sorbonne Université

180 MEMS microphones
(Invensense ICS 43-434)

→ **ordre $y = 90$**

En harmoniques sphériques
 $y = \text{floor}(n/2)$ [Rafaely 2015],
avec n microphones



Mesures - setup

Microphones
(positionnés avec
pointeur laser)

Structure circulaire (3m)
autour du participant

Participant sur une
chaise, fixe une cible
droit devant



Mesures – phonèmes enregistrés

- Participant 1 (voix parlée)

18 phonèmes, intensité vocale et F0 constants

10 voyelles: [a], [ɛ], [i], [o], [u], [y], [ø], [ɔ̃], [ɑ̃], [ɛ̃]

8 consonnes: [f], [s], [ʃ], [v], [z], [ʒ], [m], [n]

- Participant 2 (voix chantée)

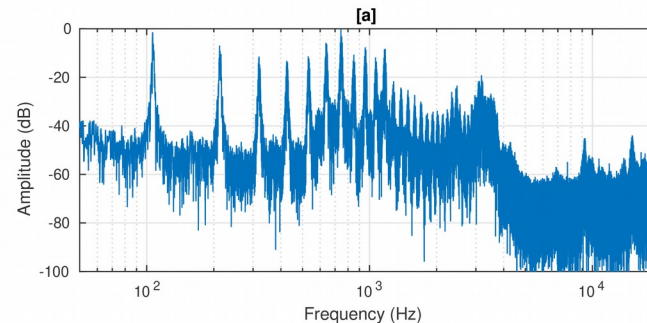
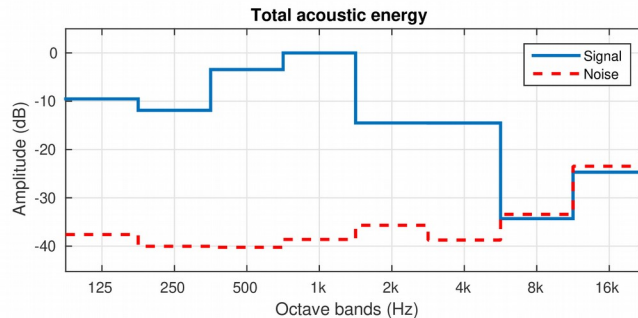
6 niveaux d'intensité vocale, F0 constant, phonème [ɛ]

a PATTE	a PÂTE	ã ANCRE	b BALLON	ɔ̃ PORTE	ɔ̃ MONTRE	d DATE	e ÉTÉ	ə FENÊTRE
ɛ MER	ɛ̃ BRIN	f FACE	g GARE	y LUI	i AMI	j AIL	k CARPE	l LA
m MAMAN	n NON	ɲ AGNEAU	o SEAU	ø JEU	œ FLEUR	œ̃ BRUN	p PETIT	ʁ RARE
s SOURIS	ʃ CHAT	t TORDU	u COUP	v VOIR	w FOUET	y NU	z ZÈBRE	ʒ JOURNAL

Mesures – voyelles [a], [i], [o]

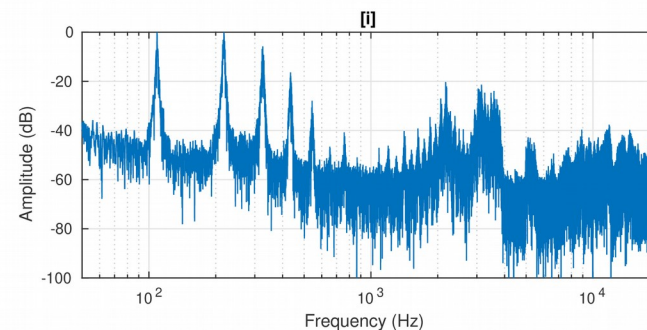
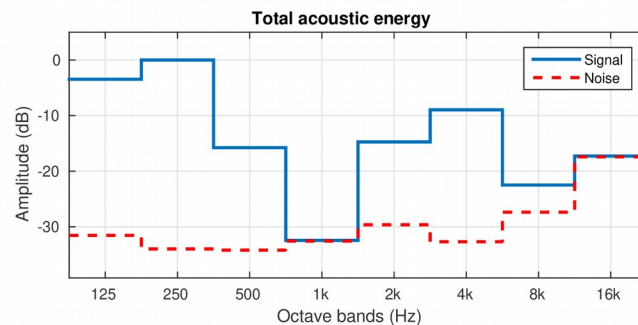
Participant 1:
Voix parlée

[a]

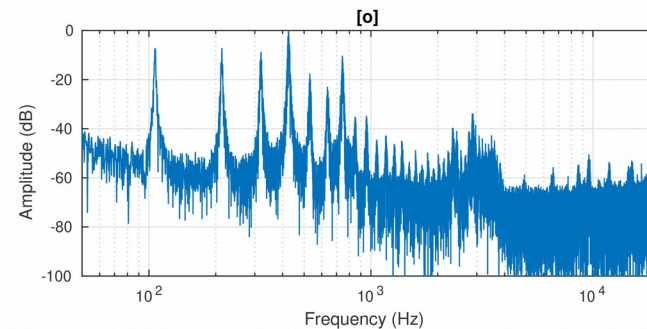
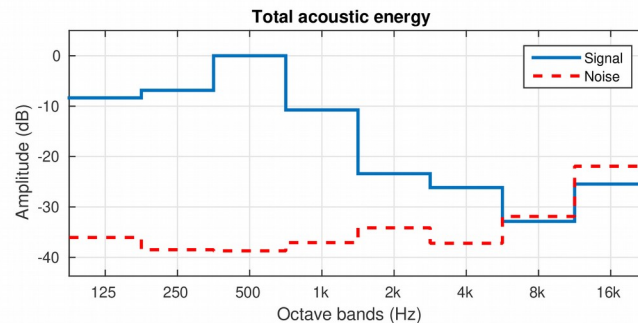


F0 = 107 Hz

[i]



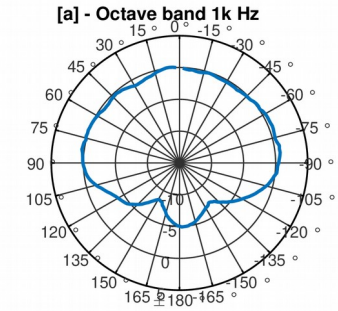
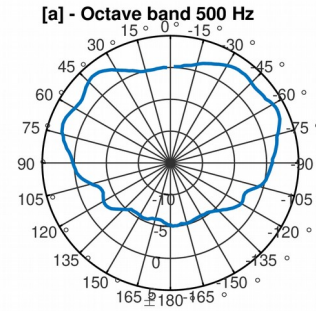
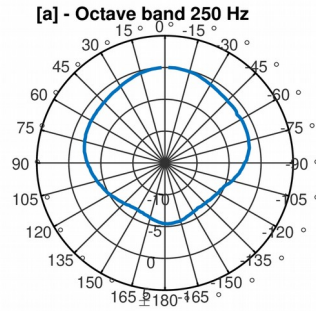
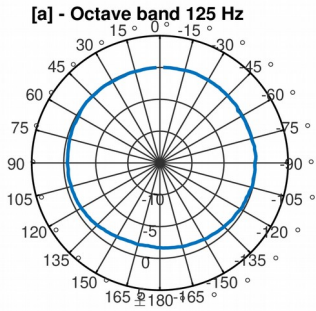
[o]



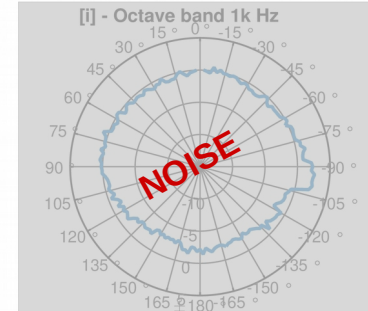
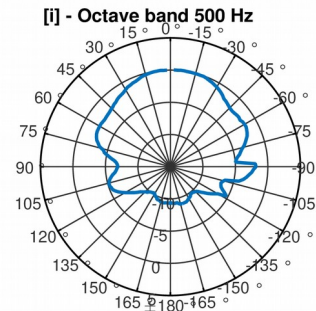
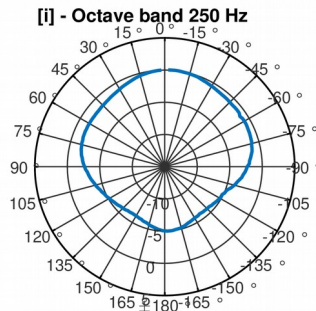
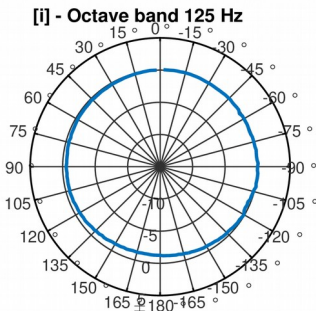
Mesures – voyelles [a], [i], [o]

[125 Hz – 1 kHz]

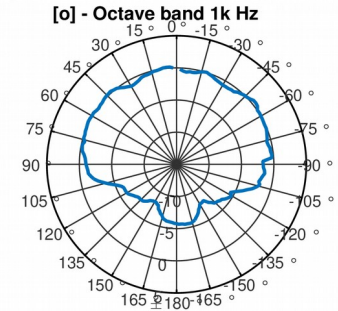
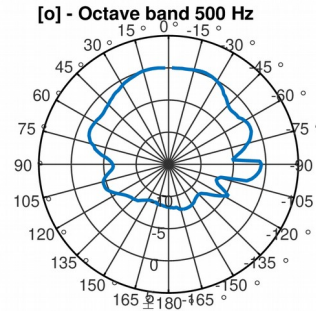
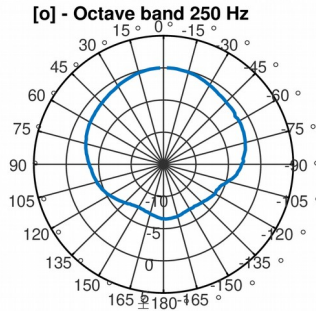
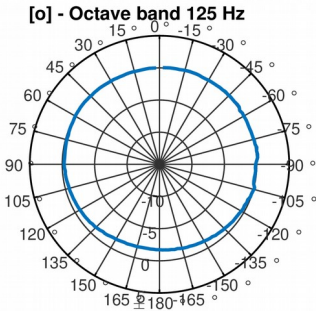
[a]



[i]



[o]



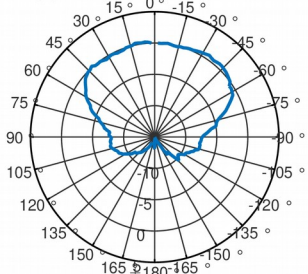
→ Différences principales à 500 Hz et au-dessus

Mesures – voyelles [a], [i], [o]

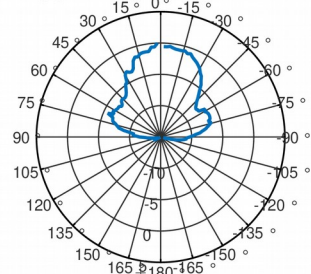
[2 kHz – 16 kHz]

[a]

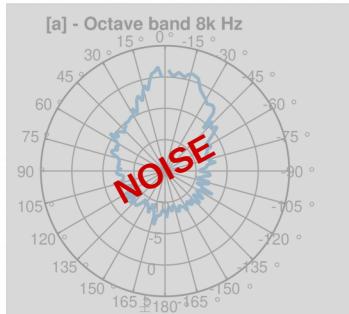
[a] - Octave band 2k Hz



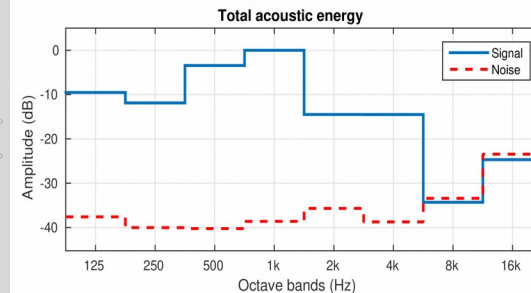
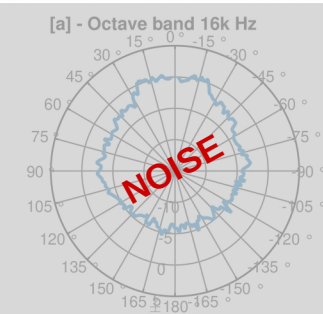
[a] - Octave band 4k Hz



[a] - Octave band 8k Hz

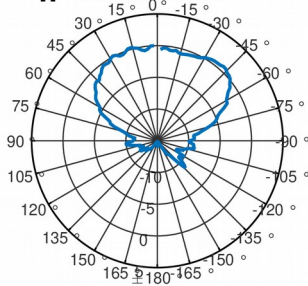


[a] - Octave band 16k Hz

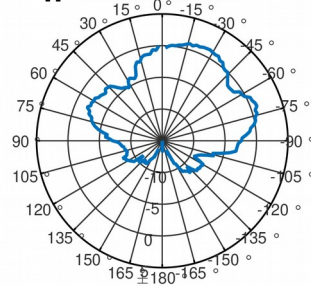


[i]

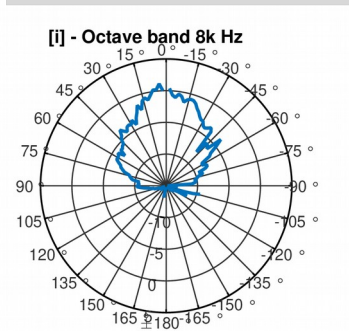
[i] - Octave band 2k Hz



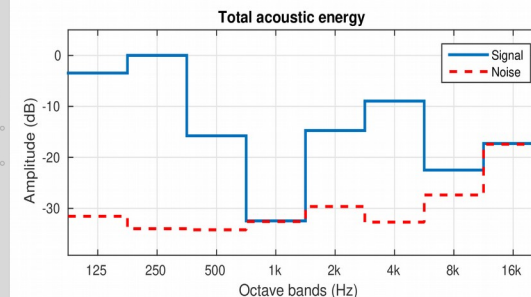
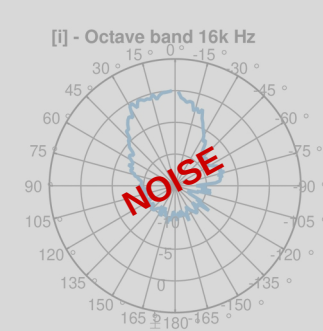
[i] - Octave band 4k Hz



[i] - Octave band 8k Hz

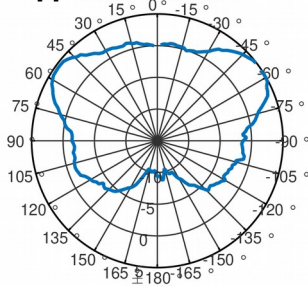


[i] - Octave band 16k Hz

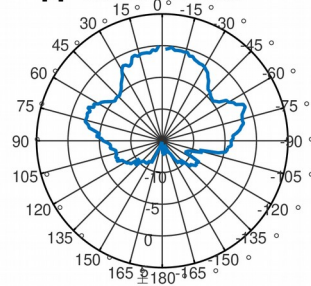


[o]

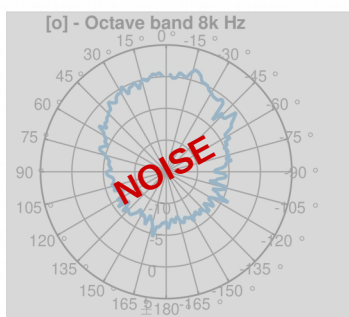
[o] - Octave band 2k Hz



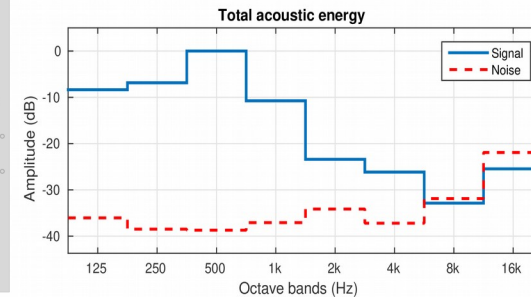
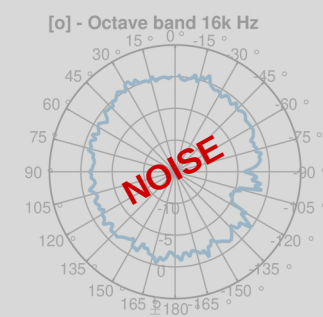
[o] - Octave band 4k Hz



[o] - Octave band 8k Hz



[o] - Octave band 16k Hz

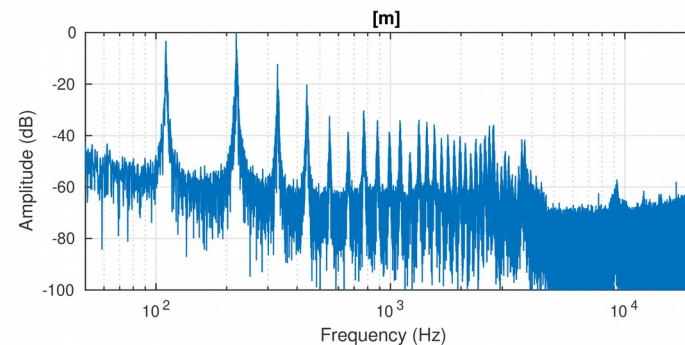
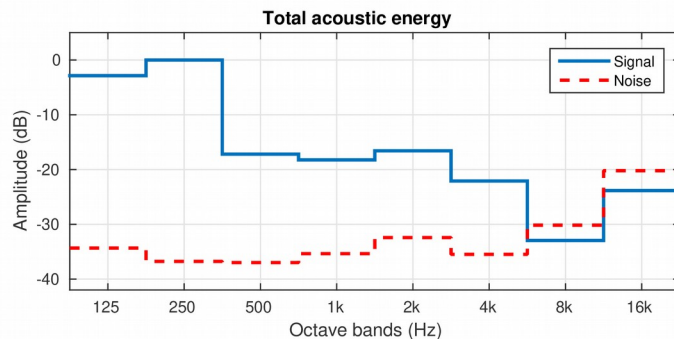


→ Très peu d'énergie à 8 kHz et 16 kHz

Mesures – consonnes [m], [v], [s]

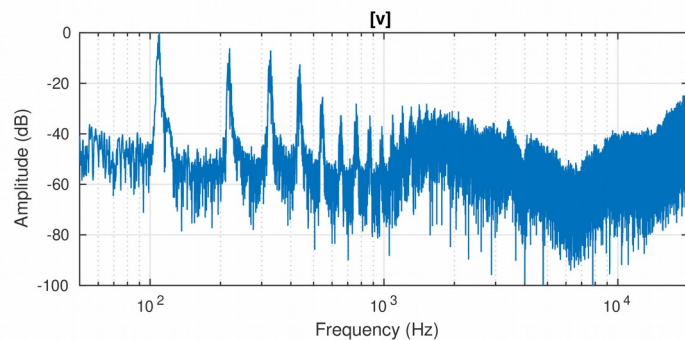
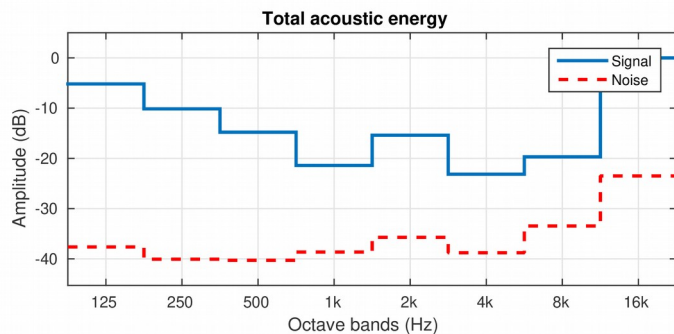
Nasale

[m]



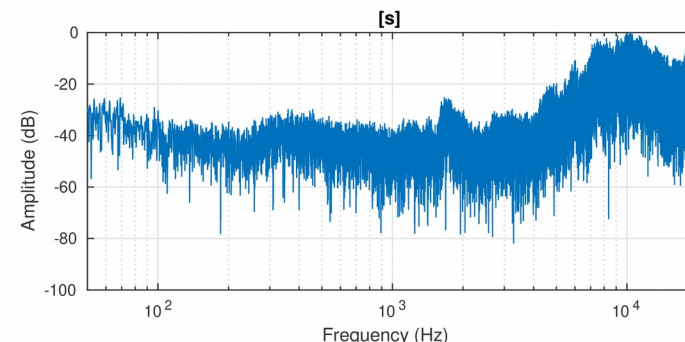
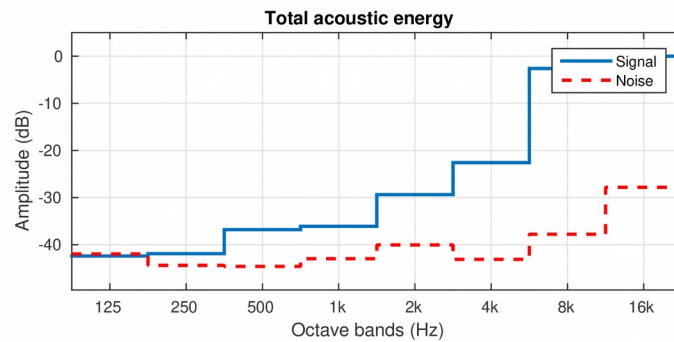
Fricative
voisée

[v]



Fricative
Non voisée

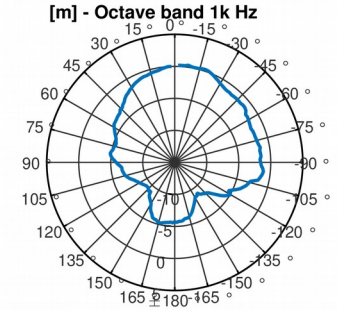
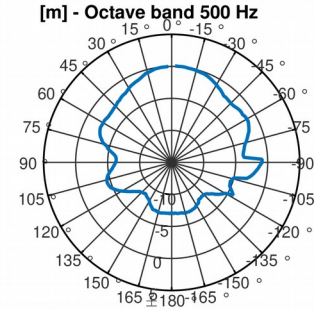
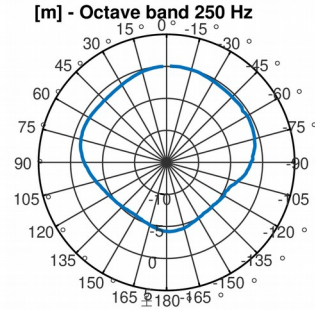
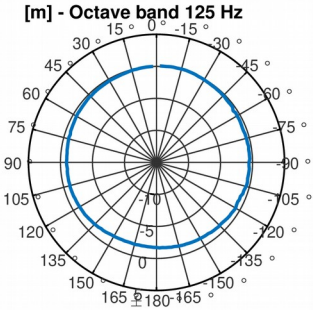
[s]



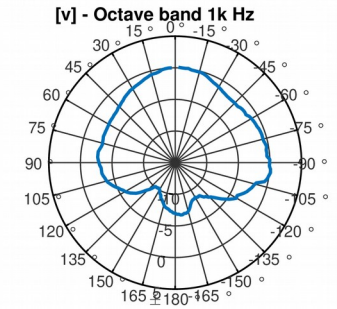
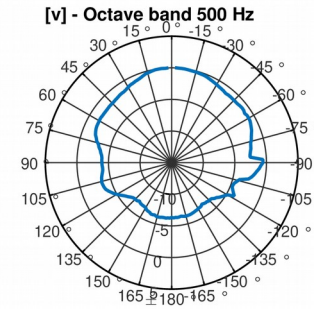
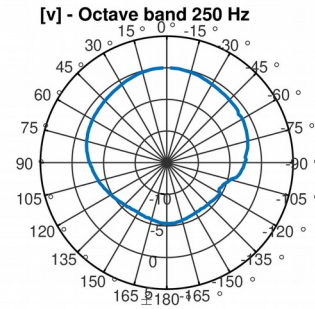
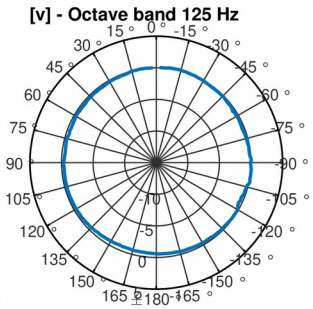
Mesures – consonnes [m], [v], [s]

Basse fréquence [125 Hz – 1 kHz]

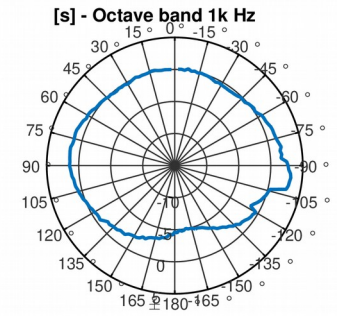
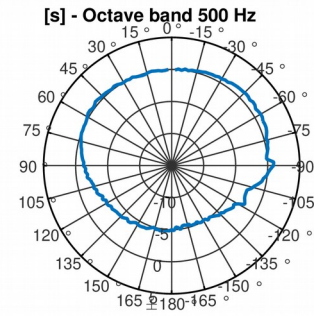
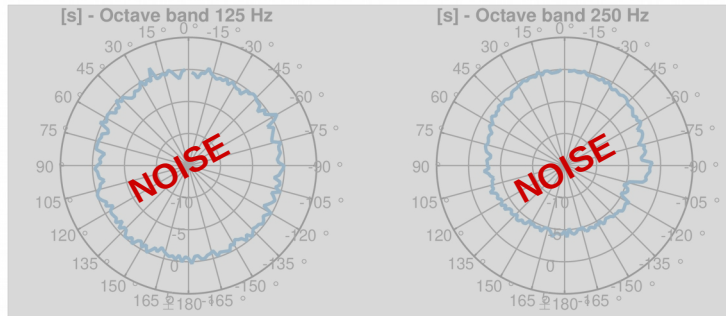
Nasale [m]



Fricative voisée [v]



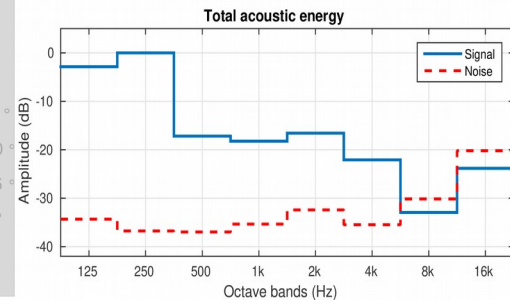
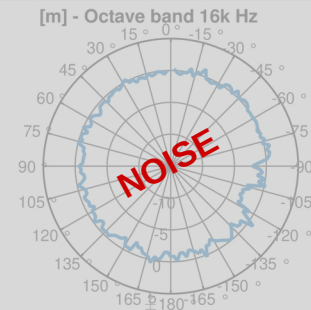
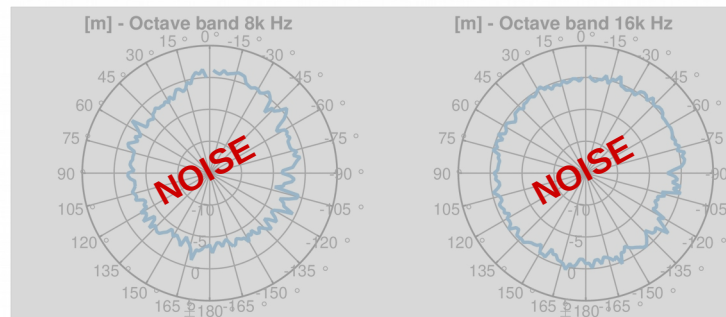
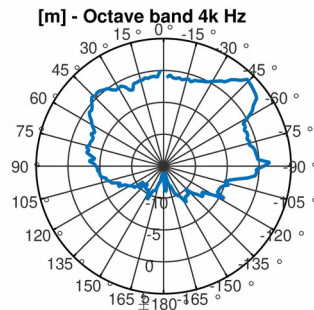
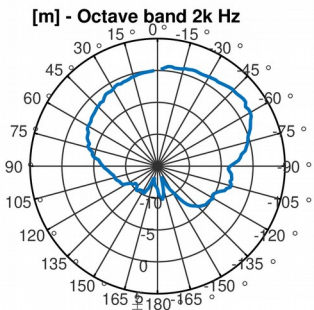
Fricative Non voisée [s]



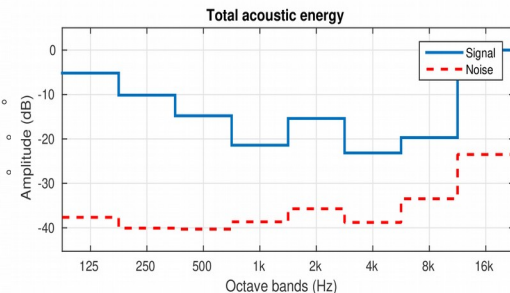
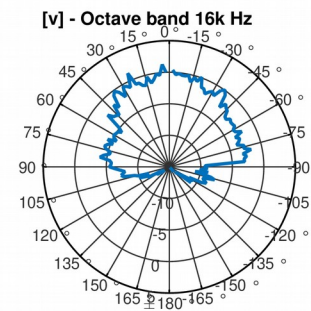
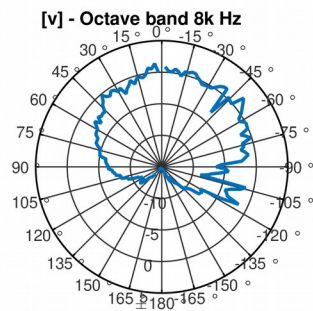
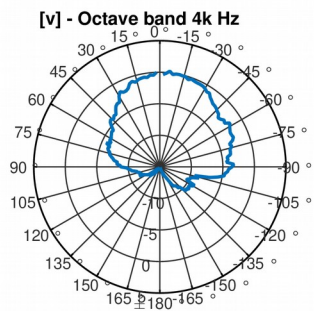
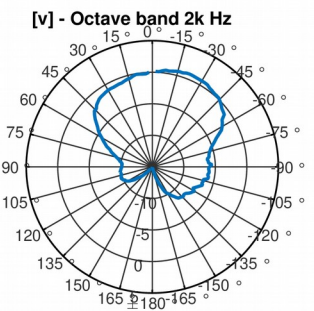
Mesures – consonnes [m], [v], [s]

Haute fréquence [2 kHz – 16 kHz]

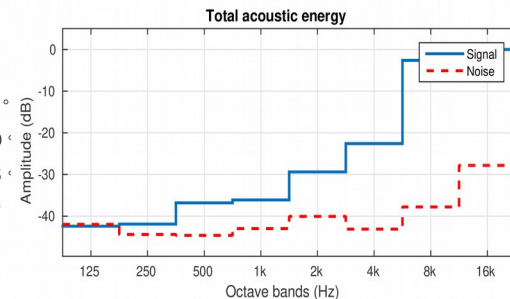
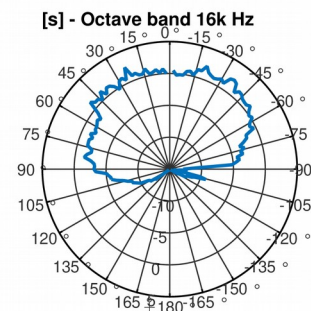
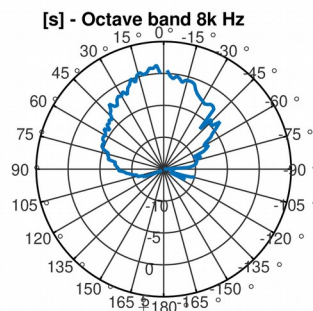
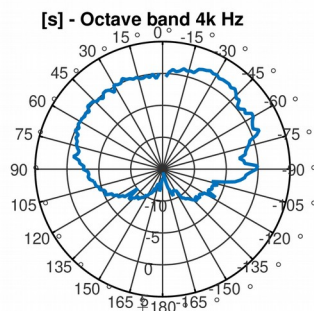
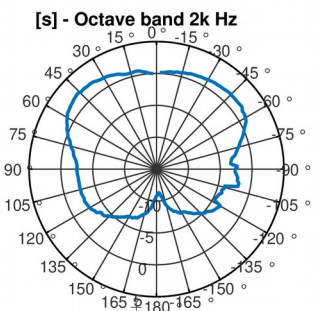
[m]



[v]



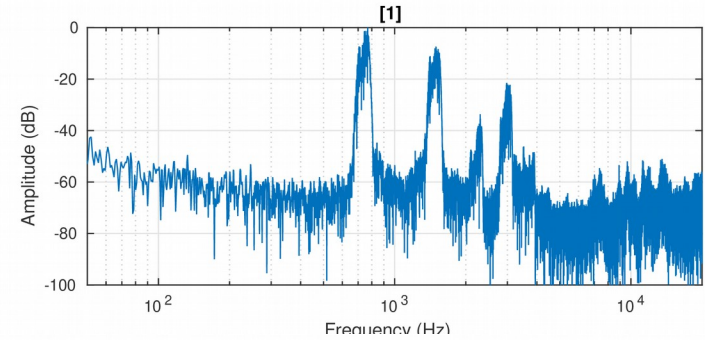
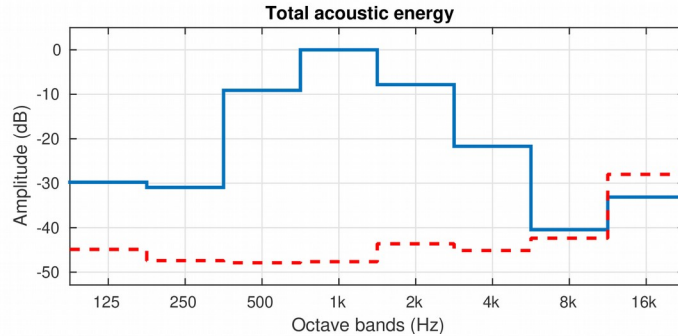
[s]



Mesures – intensité vocale

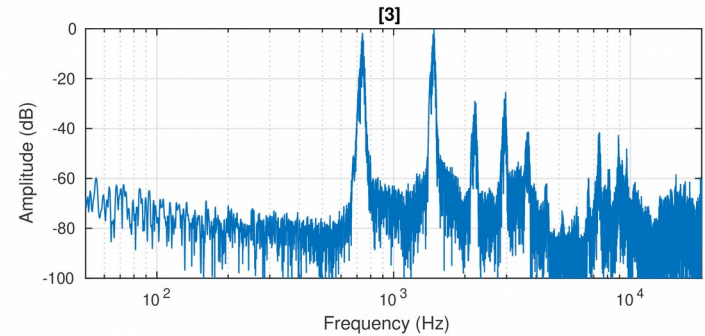
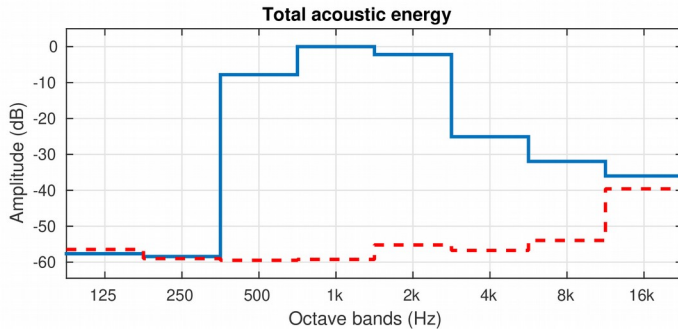
Participant 2:
Voix chantée

Intensité vocale : basse

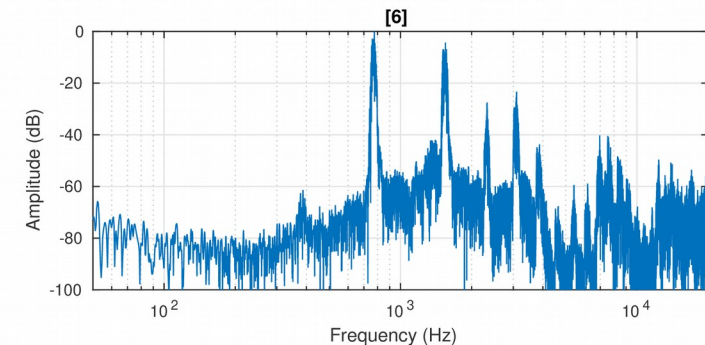
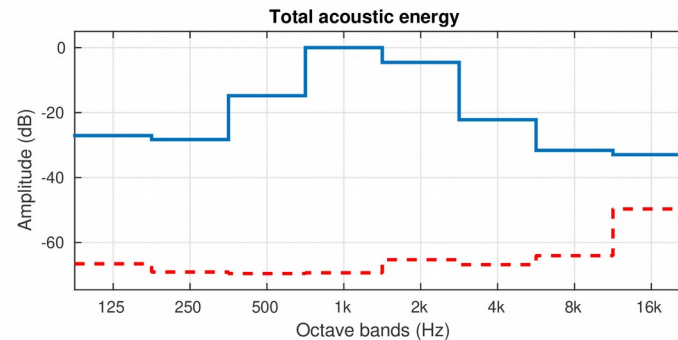


F0 = 750 Hz

Moyenne



Haute

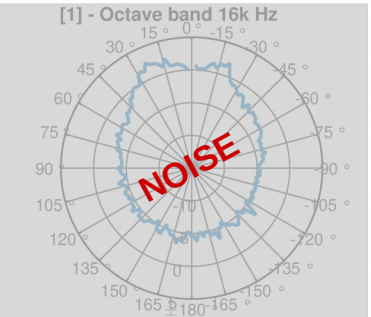
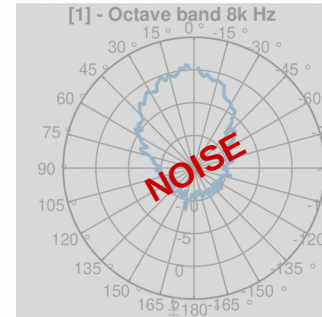
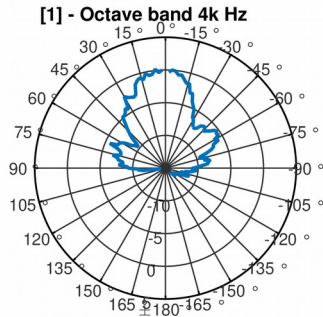
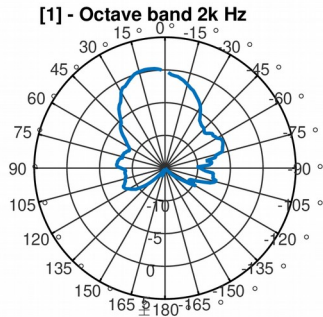


Mesures – intensité vocale

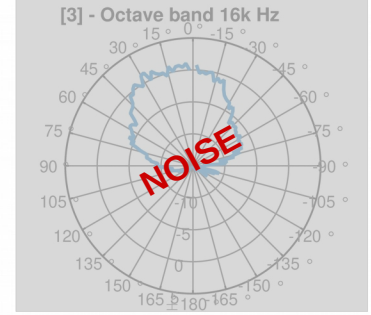
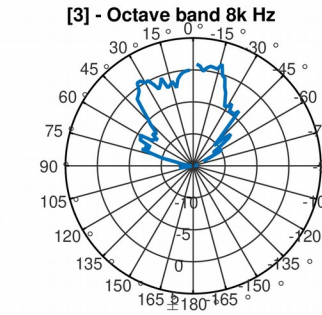
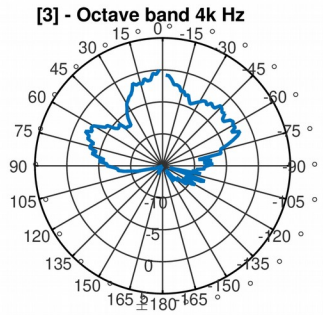
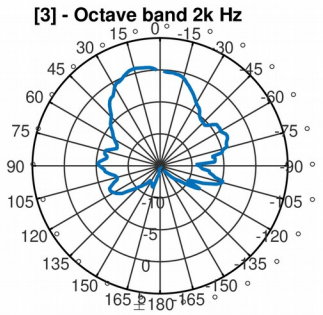
[2 kHz – 16 kHz]

Intensité vocale:

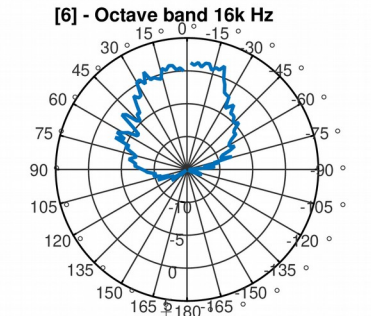
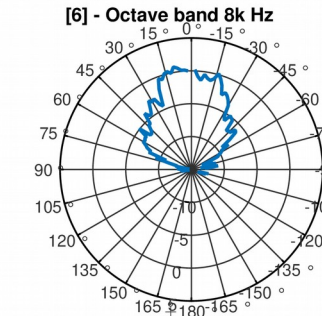
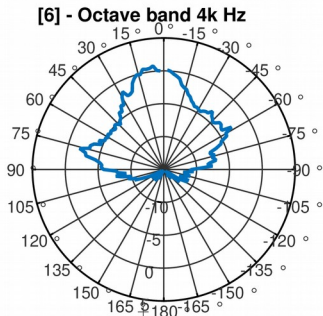
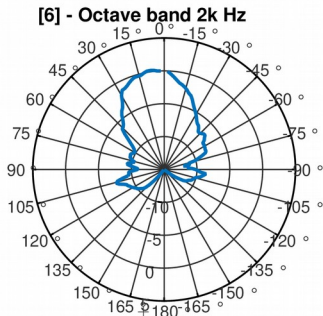
Basse



Moyenne



Haute



→ Besoin d'une méthode de caractérisation plus fine

Conclusion

- Système de mesures opérationnel en 2D
- Mesure du champ sonore jusqu'à 20 kHz, malgré le niveau de bruit élevé aux hautes fréquences
- Haute densité spatiale (résolution angulaire 2°) dans le plan horizontal
- Phonèmes : variations de directivité observées sur spectre audible. Cohérent avec études précédentes [Katz et al, 2006 ; Pörschmann & Arend, 2021]
- Intensité vocale : effets modérés, nécessite une caractérisation plus fine

Prochaines étapes

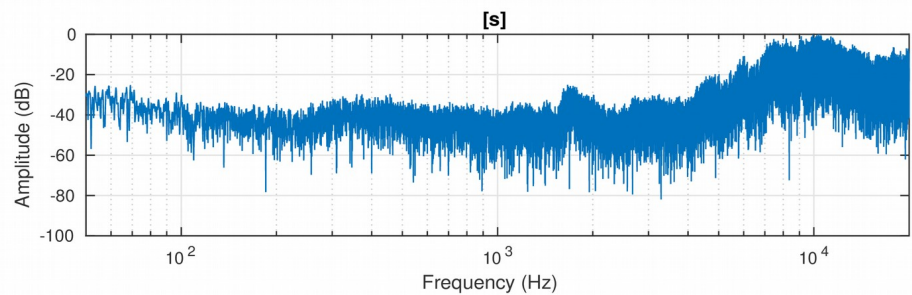
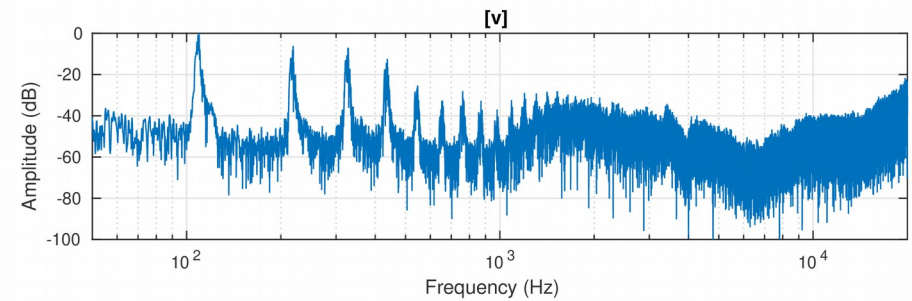
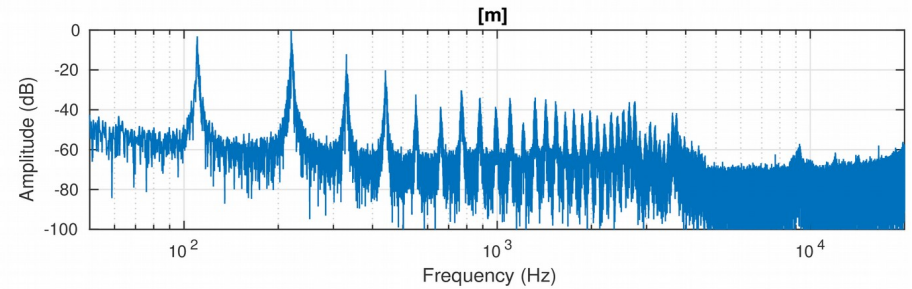
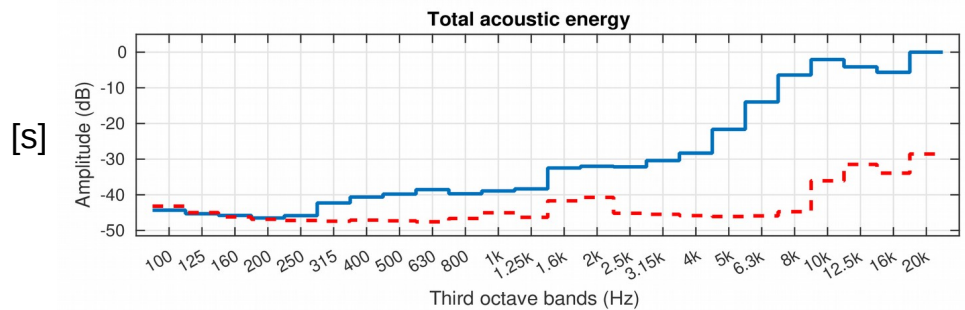
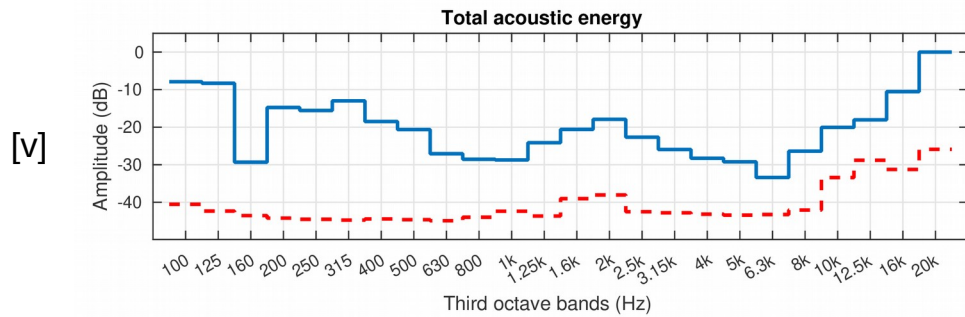
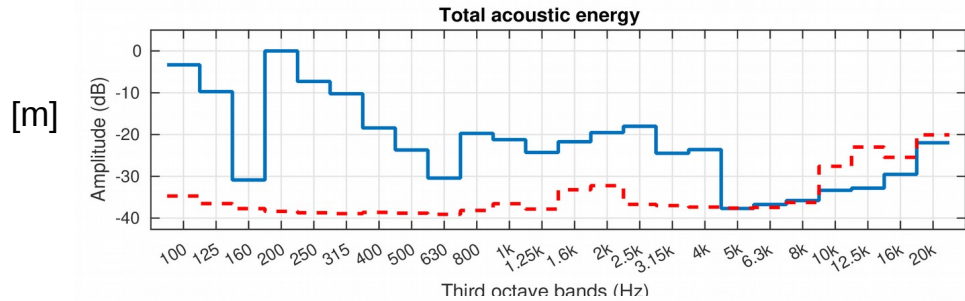
- Méthode de caractérisation de la directivité plus fine que Directivity Index (1 seule direction), e.g. avec coefficients SH
- Variabilité des phonèmes dans phrases entières (pas phonèmes isolés)?
- Construction d'une antenne de micros 3D sphérique
- Expériences perceptives : patterns complexes de directivité entendus par locuteur et auditeur?

Sources de financement

- DFG project “**Spatial characteristics of the human voice**”, 2021-2024
Project No. 457863315
- European Horizon 2020 project “**Sonicom**”, 2020-2025
Agreement No. 101017743
- Projet ANR “**RAYOVOX**”, 2021-2025
Project No. ANR-21-CE42-0017-01



Annexe – consonnes [m], [v], [s] en 1/3 octave



Annex – consonants [m], [v], [s] in 1/3 octave bands

