Objective Speech Quality Estimation of In-Ear Microphone Speech

João F. Santos, Rachel E. Bouserhal,
Jérémie Voix, Tiago H. Falk

INRS
UNIVERSITÉ DE RECHERCHE

ÉTS
Le génie pour l'industrie

CIRMMT

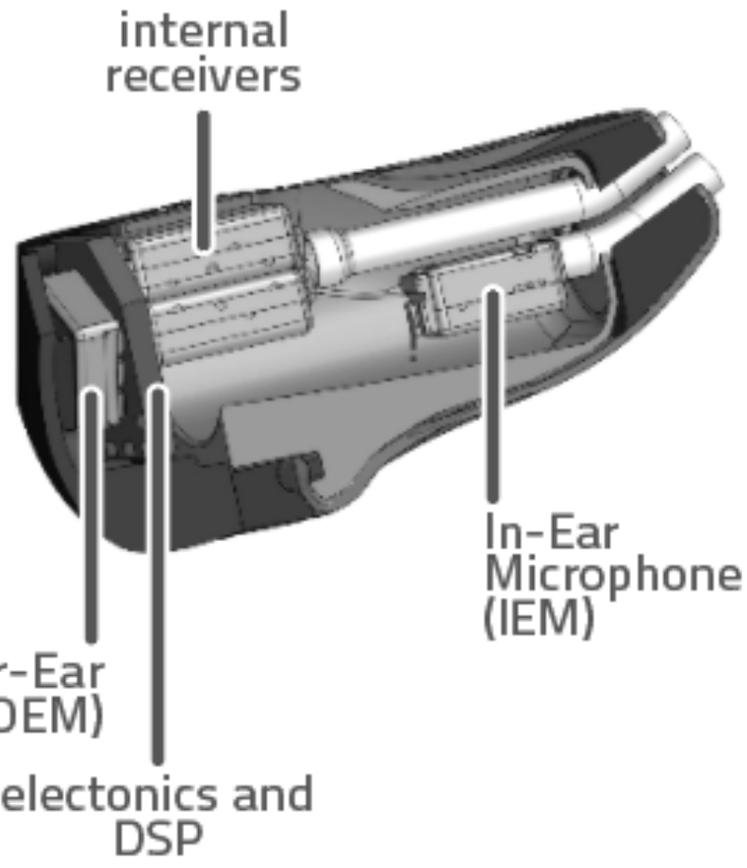
Introduction

- Speech captured from in-ear microphones (IEM) in extremely noisy environments maintains a high SNR
- Limited bandwidth due to occlusion, so speech enhancement is required to obtain more natural speech
- Our goal: objectively estimate speech quality of (noisy/enhanced) IEM speech

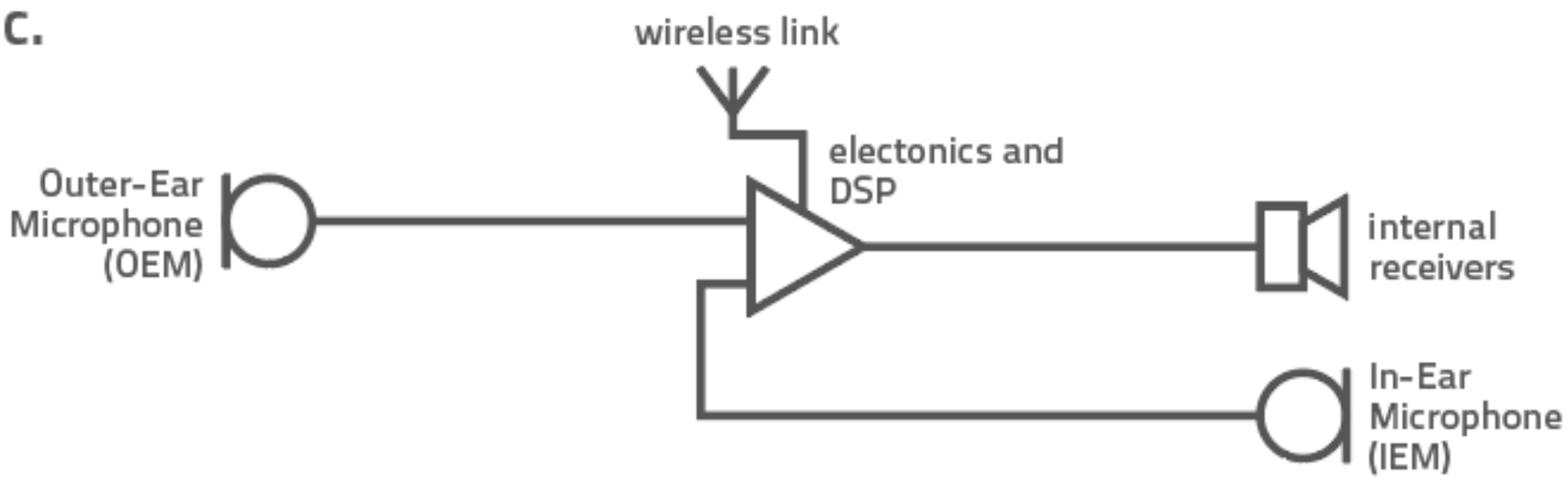
a.


inflatable
earpiece
wireless link

b.


internal
receivers
In-Ear
Microphone
(IEM)
Outer-Ear
Microphone (OEM)
electonics and
DSP

C.


wireless link
electonics and
DSP
Outer-Ear
Microphone
(OEM)
internal
receivers
In-Ear
Microphone
(IEM)

In-ear speech quality dataset

- Speech recorded in an audiometric booth, using both the in-ear device and an external Zoom®H4n

- Simulated residual noise “leaks" by attenuating signal mixed with factory noise at -5 dB SNR using measured attenuation of the device

- Noisy signals were denoised using an adaptive nLMS algorithm and bandwidth extended in the high frequencies

In-ear speech quality dataset

- Online MUlti Stimulus Test with Hidden Reference and Anchor (MUSHRA)

- Reference: signal recorded with external recorder

- Anchors: noisy and bandlimited and noise-corrupted version of the reference signal

- 42 participants with self-reported normal hearing took part in the test

Benchmark metrics

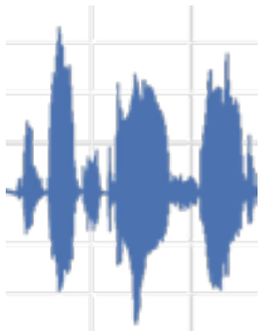
Intrusive metrics:

- PESQ
- POLQA

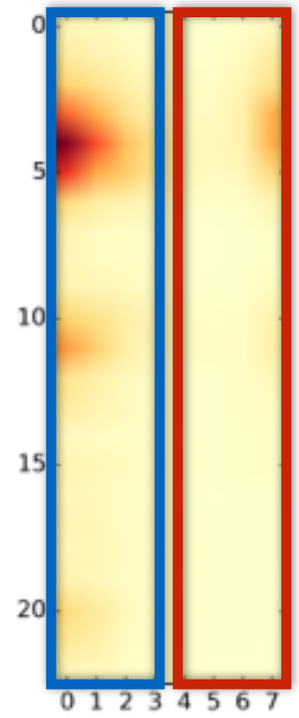
Non-intrusive metrics:

- ANIQUE+
- P.563
- SRMR

Modulation spectrum representation of speech signals

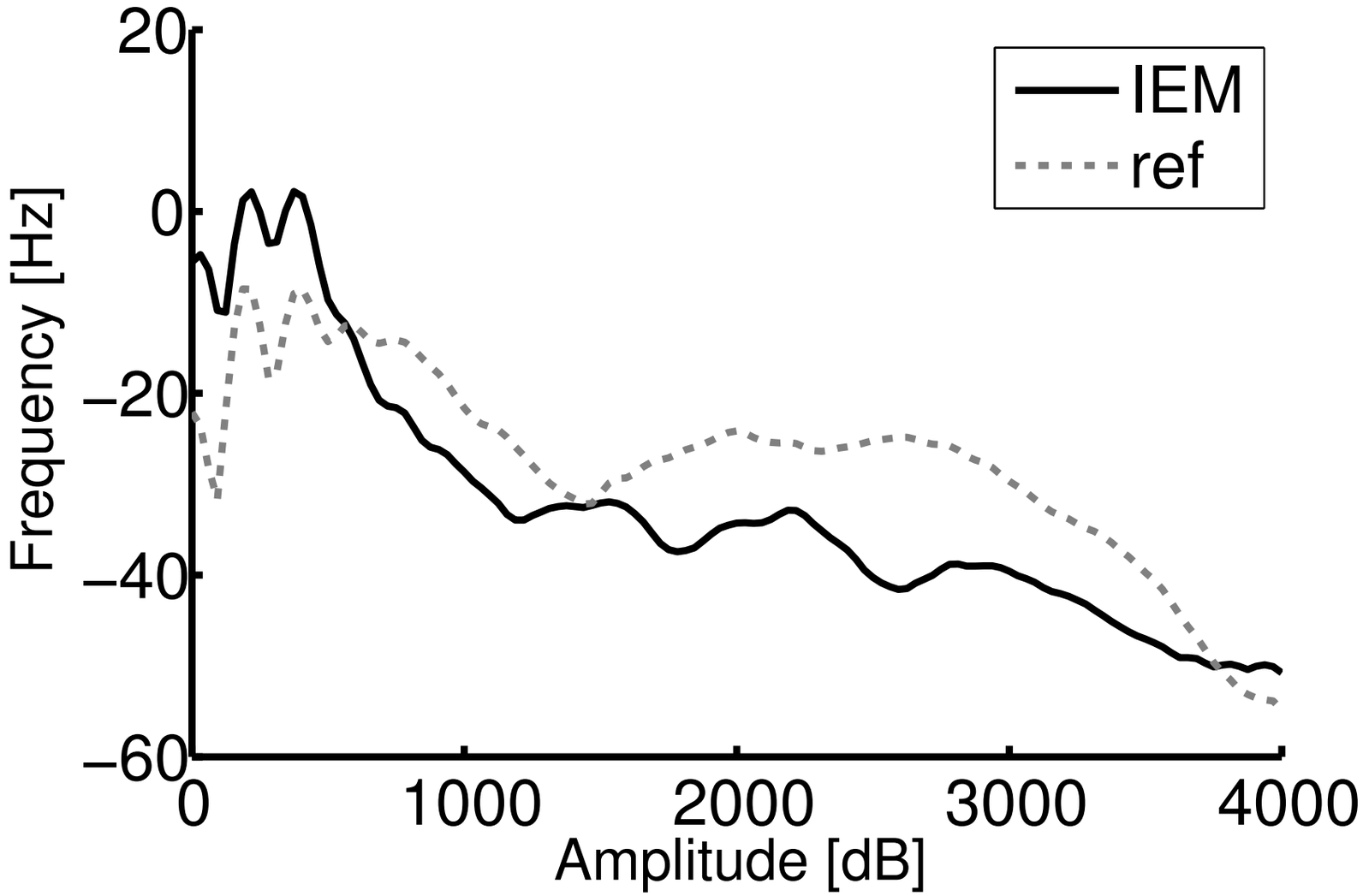


Cochlear filterbank → Envelope extraction → Modulation filterbank →

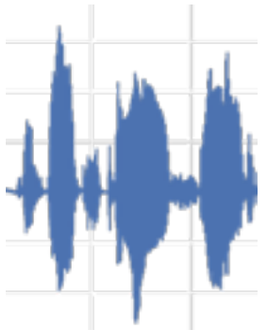


$$\text{SRMR} = \frac{\sum(\text{Energy in low MF})}{\sum(\text{Energy in high MF})}$$

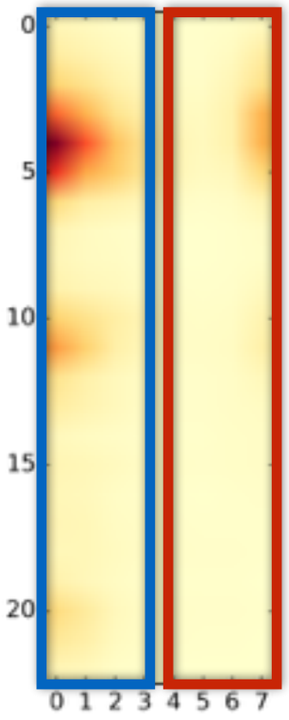
In-ear vs. reference speech



Proposed metric



Modulation analysis



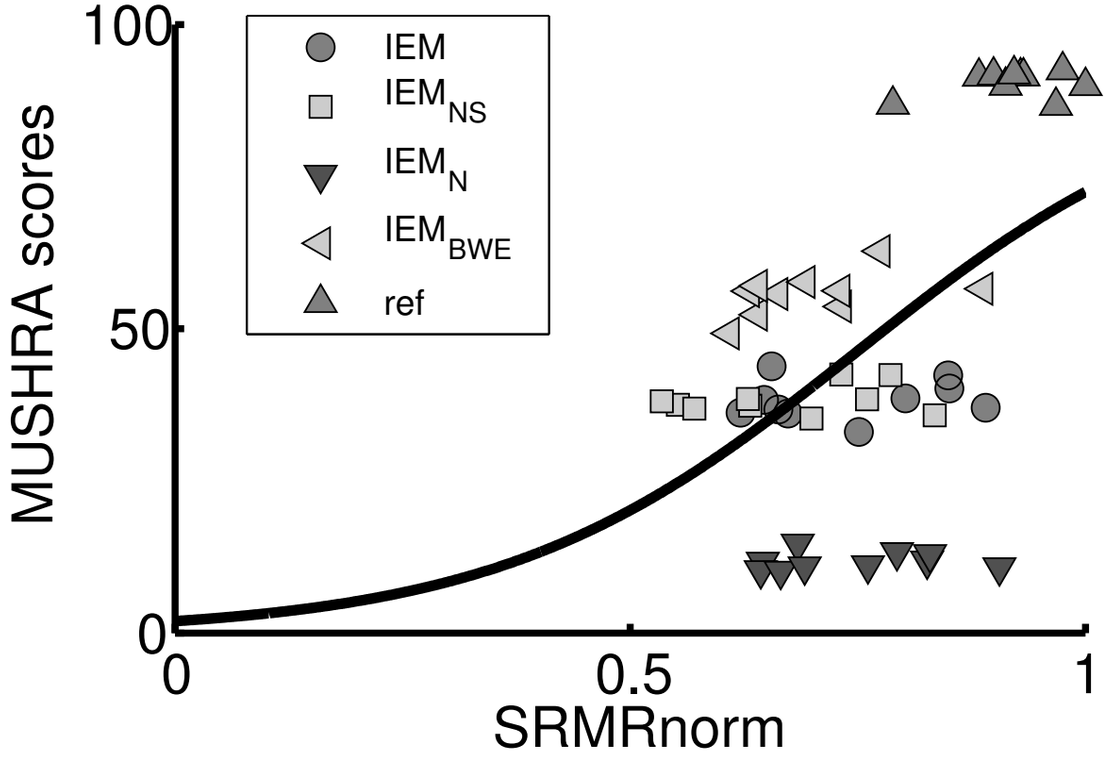
Mod. energy normalization

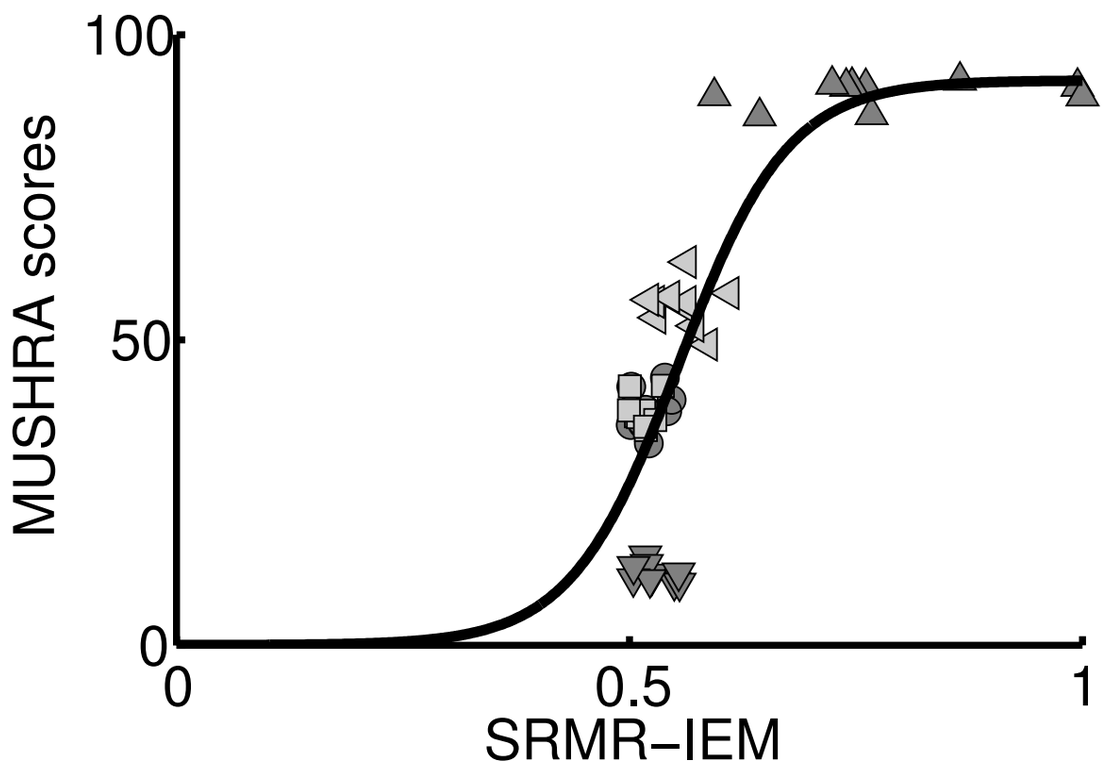
Weighting function

15 dB attenuation, $f_c < 2$ kHz
15 dB gain, $f_c > 2$ kHz

Results

Metric	ρ	ρ_{sp}	ρ_{sig}	RMSE
PESQ	0.922	0.685	0.903	11.20
POLQA	0.906	0.814	0.894	11.65
P.563	0.561	0.522	0.585	21.11
ANIQUE+	0.531	0.502	0.535	21.98
SRMRnorm (30 dB)	0.536	0.408	0.528	22.10
SRMR-IEM (15 dB)	0.811	0.692	0.867	12.98
SRMR-IEM (30 dB)	0.727	0.533	0.728	17.84
SRMR-IEM (60 dB)	0.530	0.384	0.523	22.18


100
50
0
0
0.5
1
MUSHRA scores
SRMRnorm
IEM
IEM_NS
IEM_N
IEM_BWE
ref


100
50
0
0
0.5
1
MUSHRA scores
SRMR–IEM

Discussion

- SRMR-IEM does not discriminate the IEM noisy scenario from the other IEM scenarios: modulation spectra are too similar

- Study limitations:

  - Weighting based on speech by a single female speaker

  - Limited number of scenarios

Conclusions

- Most non-intrusive metrics showed poor performance compared to intrusive metrics
- SRMR-IEM significantly reduces this performance gap
- Future work: evaluate the proposed metric on a larger dataset