Sonification & speech: Recent research on auditory feedback based on human movement

Benjamin O'Brien, Ph.D
http://benjaminobrien.net

Outline

- Introduction
- Sonification * Cycling
- Pathological Speech Applications
  - Spectrotemporal modulation
  - Hypoxic modeling
- Reflections

Introduction : Who am I?

with Big Band du CNRS, Marseille (2022)

- University of Virginia — BA Mathematics (2006)
- Mills — MA Composition (2009)
- UF University of Florida — PhD Composition (2015)
- Aix*Marseille université, Initiative d'excellence — M2MPI – acoustics (2017)
- Institut des Sciences du Mouvement Etienne Jules Marey — Post-doctorate (2017-2019)
- UMR 7309 • CNRS • AMU Parole et Langage — Post-doctorate (2020-2022)
- Laboratoire Informatique d'Avignon, Université d'Avignon et des Pays de Vaucluse — Post-doctorate (2022-Present)

Introduction: Example

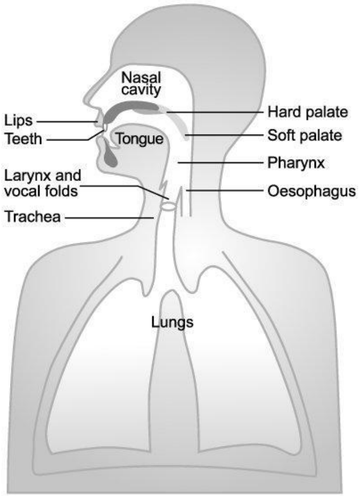
Relationship between movement and sound

Introduction: Natural & artificial auditory feedback

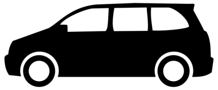
"Sound provides information about an interaction of materials at a location in an environment" – Gaver (1993)

Natural

- Wind (mistral!)
- Cicadae
- Speech*



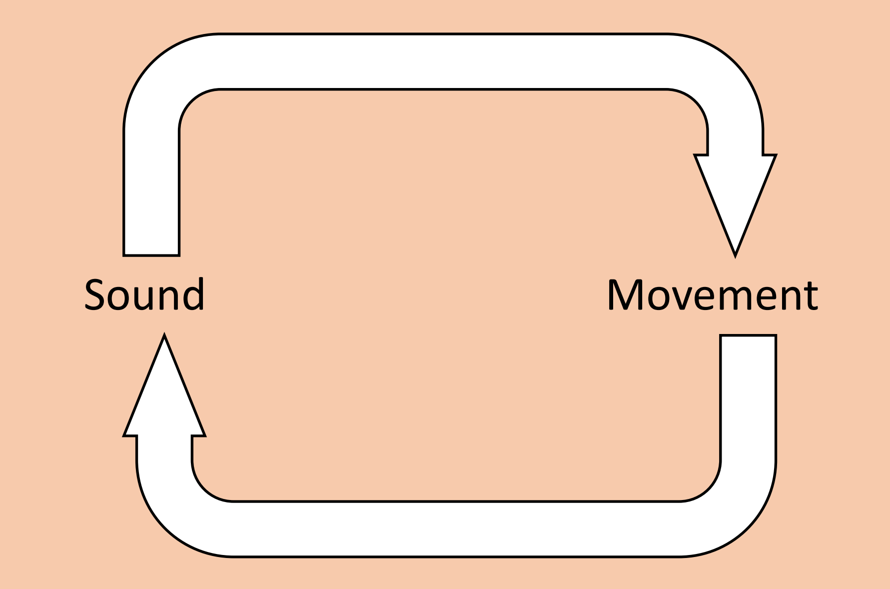
Artificial

- Transportation 
- Thermal
- Medical

* To be discussed later . . .

Introduction: Artificial "sonification"

"…rendering sound in response to data" – Hermann et al. (2011)



Comparison with visual feedback

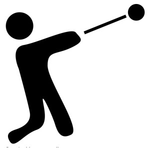
- ✔ Sound is **more temporally accurate** than vision (Hirsh & Watson, 1996; Murgia et al., 2017)
- ✔ **Less consuming** on **attentional** resources
- ✔ **Portable** with the use of modern technology (Secoli et al., 2011)

Introduction: Sonification studies

Outside & ....

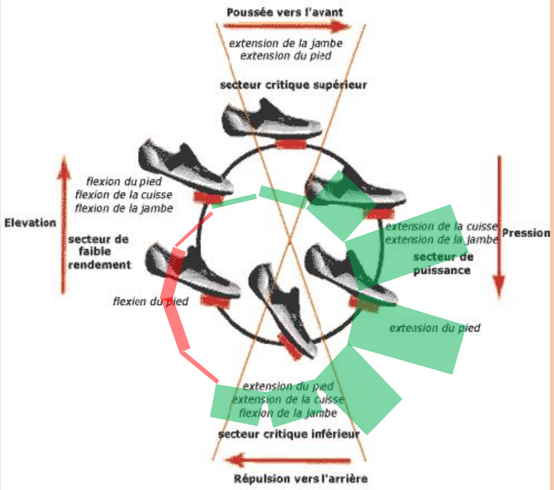
- Auditory cues effective in informing **spatio-temporal** aspects in drawing tasks (Thoret et al., 2014; 2016)
- Kinematic sonification during learning of new **handwriting** task improved **speed** & **fluency** of **new skill** (Danna et al., 2014)

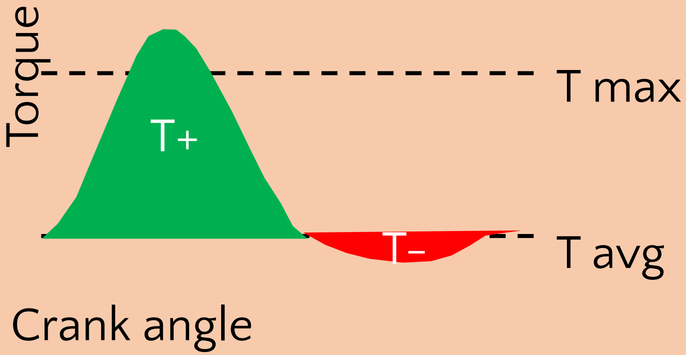
... inside sports domain



- Listening to acoustic recording of best **hammer throw** performance enhanced performance (Agostini et al., 2004)
- 4-D sonification enhanced motor learning of **rowing motion** skill (Effenberg et al., 2016)

Sonification: Pedaling gesture





Useful

Resistant

Bertucci and Grappe (2009)

Henke (1998)

scientific reports

OPEN

Online sonification improves cycling performance through kinematic and muscular reorganisations

Benjamin O'Brien, Romain Hardouin, Guillaume Rao, Denis Bertin & Christophe Bourdin

- ✔ Pedal stroke **rarely optimal**
- ✔ **Visual feedback** could help, but incompatible with outdoor cycling (Biddio et al., 2012)

Aix*Marseille université Initiative d'excellence · CNRS · ANR · SATT Sud-Est · Institut des Sciences du Mouvement Etienne Jules Marey · TechnoSport

Sonification: Previous development

Goal: Can sonification be used to enhance performance? (Vidal et al., 2019)

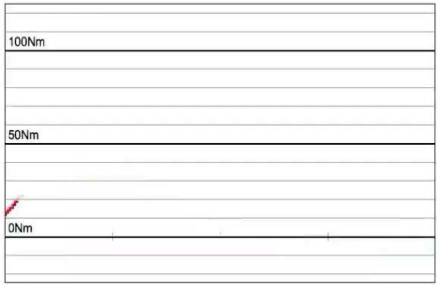
✔ Sonification of **stereo** most **efficient**

Method

- Torque efficiency (TE)

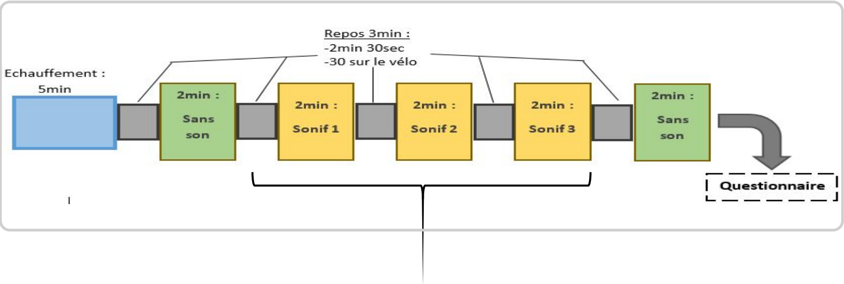
$$TE = 100 * \frac{\tau^+}{\tau^+ + \tau^-}$$

- ***Squeak*** when < 0 NM

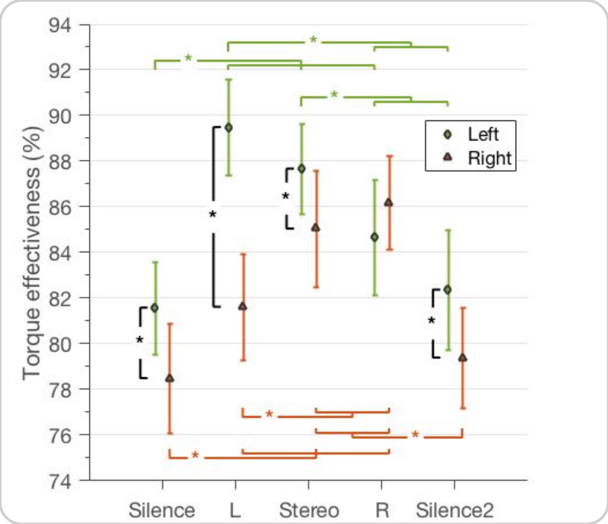


Participants

- 12 experts; 16 naïve
- Different display conditions



Results



Sonification: Effects on performance

Goal: Effects of sonification on kinematic activities & muscular activations
https://www.youtube.com/watch?v=65QHzRo74Bo

Methods

- Torque efficiency (TE)
- Joint kinematics
- Muscular activations (RL only)

Participants

- 8 experts
- 4 sessions (6 min):
  - Silence
  - Left
  - Right
  - Stereo

Experimental setup

20” sonif

40” no sonif

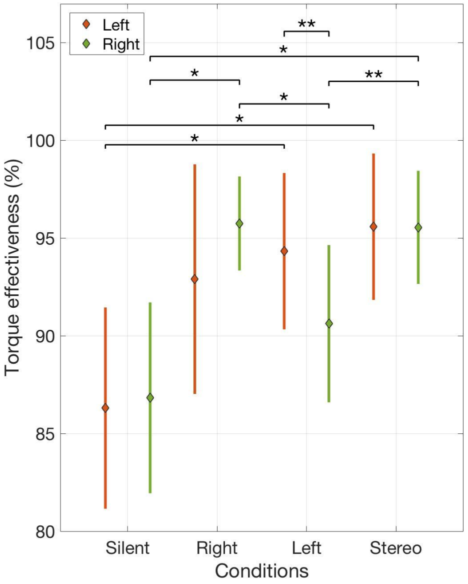
1’ warm up

1’ cool down

Sonification: Results

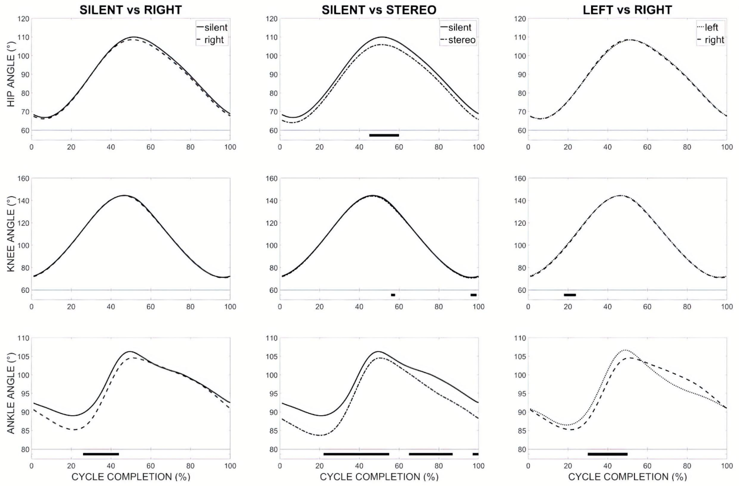
Are these limitations or **adaptations**?

TE



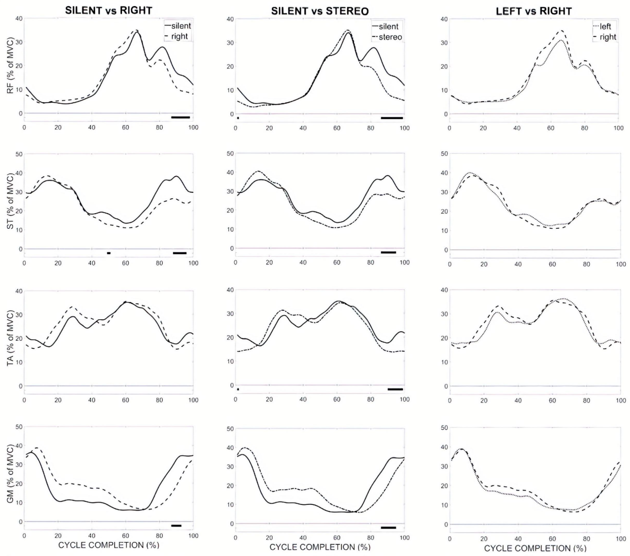
✔ TE performance enhanced with **sonification**

Kinematic



✔ Sonification **locked ankle**, ascend chain to **knee**, little effect on hip

Muscular



✔ Sonification **limited** movement → less need to **activate** muscles

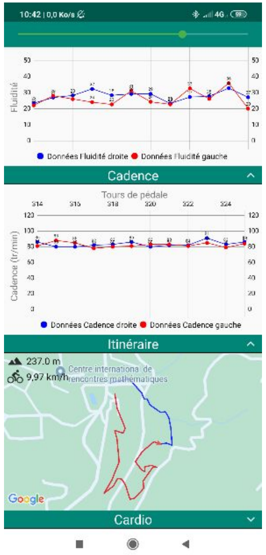
Aix*Marseille université Initiative d'excellence
cnrs
ANR
SATT SUD-EST
INSTITUT DES SCIENCES DU MOUVEMENT ETIENNE JULES MAREY
TechnoSport

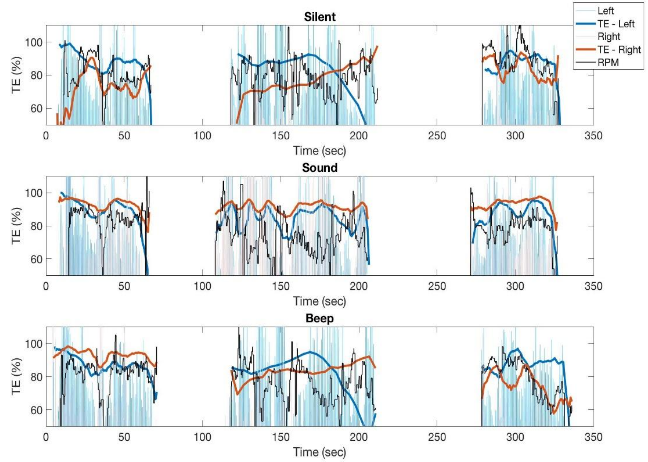
Sonification: Application

Testing to start in April 2023

headphones

phone (mount)





Speech: Signal analysis

Speech signals **transmit** a wealth of information pertaining to speakers, including ***physiological*** and cultural markers (Schweinberg et al., 2014; Benzeghiba et al., 2007)

Perceptual

- Intelligibility
- Prosody → emotion
- Speech rate

Acoustic

- F0 & formants
- MFCC
- Modulation

Speech: Pathological speech → STM (1/2)

Problem Metrics used by clinicians to evaluate speech impairment → time consuming

Goal Identify acoustic correlates → potential for automation

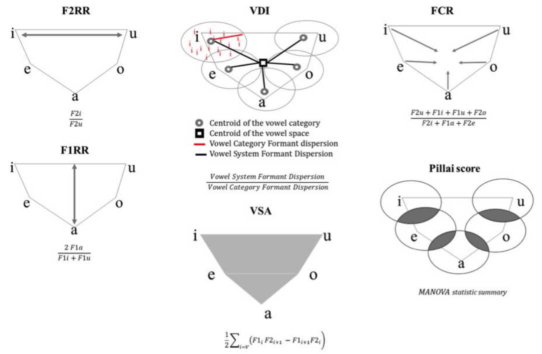
Corpus 3 datasets* → 1 training; 2 testing (HNC, Park)

spectrotemporal modulation extraction:

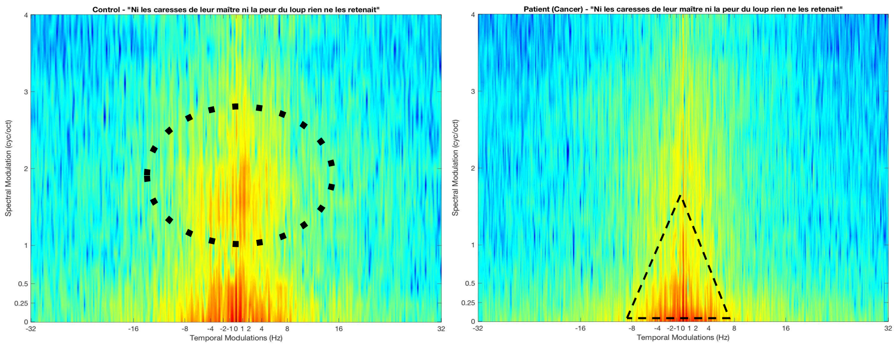
apply FFT to spectrogram to transfer time-frequency to modulation domain

Why?

Vowel quality metrics



spectrotemporal modulation domain



Temporal modulations (x-axis)
- Prosody (1-2Hz)
- Syllabic (4-8 Hz)
- Articulation (16-32Hz)

Spectral modulations (y-axis)
- Vowels (0-3 cyc/octave)
- Perceptual threshold

* see bib for details

Speech: Pathological speech → STM (2/2)

EFA TM (8-32 Hz) and SM (0.5-4 cyc/oct) correlates to vowel quality metrics (train dataset)

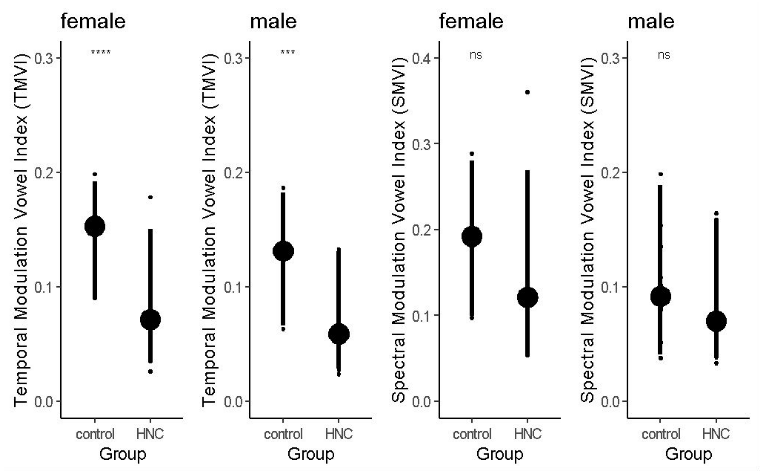
Results TM strong discriminate for HNC; not Park

Reflections TM → speech form; SM → physiological

JASA ARTICLE

Correlates of vowel clarity in the spectrotemporal modulation domain: Application to speech impairment evaluation

Anna Marczyk,[1,a] Benjamin O'Brien,[1] Pascale Tremblay,[2,b] Virginie Woisard,[3,c] and Alain Ghio[1]
[1]Aix-Marseille Université, CNRS, LPL, UMR 7309, Aix-en-Provence, France
[2]Universite Laval, Faculte de Medecine, Departement de Readaptation, Quebec City, Quebec G1V 0A6, Canada
[3]Service ORL, CHU Larrey, Toulouse, France



Model	AUC of ROC		Accuracy		Sensitivity		Specificity		Balanced Accuracy	
	M	F	M	F	M	F	M	F	M	F
MPS-based model	0.95	0.98	0.91	0.91	0.98	0.97	0.6	0.77	0.79	0.87
MPS-based vowel model	0.93	0.96	0.89	0.87	0.98	0.97	0.5	0.61	0.74	0.79
TM Vowel Index model	0.87	0.91	0.91	0.85	1	0.97	0.5	0.54	0.75	0.75
SM Vowel Index model	0.68	0.69	0.76	0.62	0.8	0.62	0.5	0.61	0.65	0.61

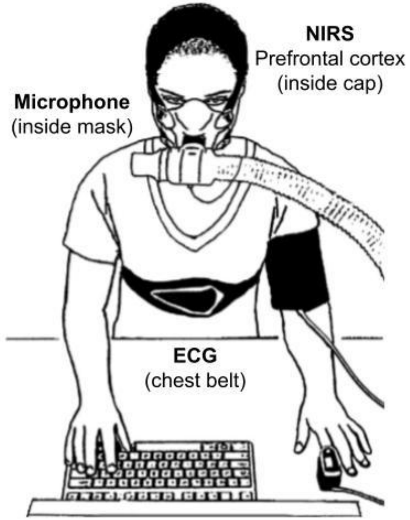
Speech: Pathological speech → Hypoxia (1/2)

Problem Hypoxia detection → slurred speech→ impaired psycho-motor skills

hypoxia: lack of oxygenation that generally occurs at high altitudes

Goal Examine effects of speech on acoustic, physio “hypoxic” features

Corpus Collected at EUROMOV in 2020 → Fresnel et al. 2021

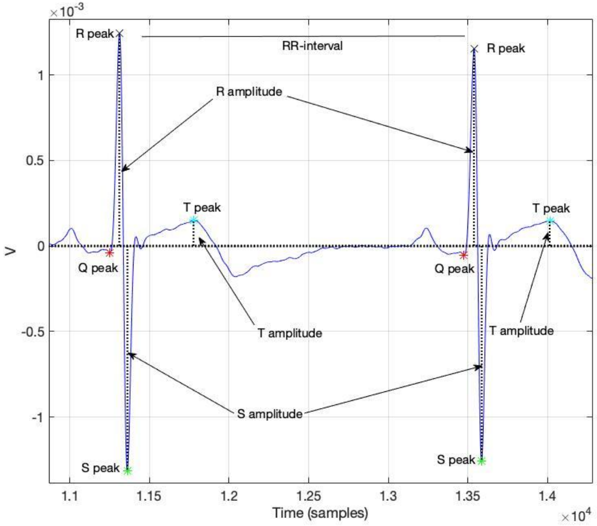


Features

Acoustic: 19 MFCC, Δ, ΔΔ

Nirs: HB tot, rSaO2

ECG: Q, R, S, T, RR-int



Speech: Pathological speech → Hypoxia (2/2)

VAD Segment into non-/speech (Speech Brain)

SLDA Identify physiological features sensitive to hypoxia

Train/Evaluate models with training dataset (SVM → 4-fold)

Reflections train (see fig); test: ~72% accuracy (a+p model)
physio > acoustic
acoustic → info about speaker → personalize?

Differentiating acoustic and physiological features in speech for hypoxia detection

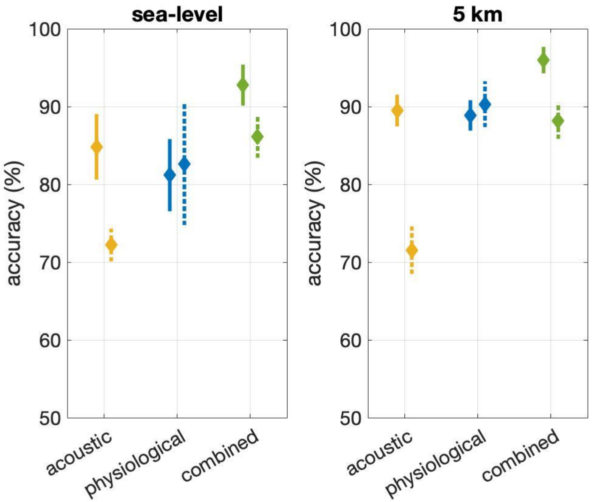
submitted to Interspeech 2023

Benjamin O'Brien[1], *Adrien Gresse*[2], *Jean-Baptiste Billaud*[2], *Guilhem Belda*[2], *Jean-François Bonastre*[1],

[1]Laboratoire Informatique d'Avignon, EA 4128, Avignon Université, France
[2]Semaxone, France

Feature	Statistic	Wilks' Lambda	F	p
RR-interval	Median	0.70	191.55	***
$rSaO_2$	Median	0.65	37.21	***
R peak	Mean	0.63	12.70	***
Q peak	Median	0.6	20.82	***
S peak	Median	0.59	9.37	**
Q peak	Std. Dev	0.58	9.00	**

training



Closing thoughts

Sonification

* personalized sonification based on physiology, perceptual capacities
* multimodal
* learning, improving, rehabilitation

Speech

* speech pathology → linked to physiological features
* spectrotemporal modulation → associated with (in)voluntary movement
* hypoxia → physio > BUT speech may provide information about speaker

Sonification & speech: Recent research on auditory feedback based on human movement

Benjamin O'Brien, Ph.D
http://benjaminobrien.net

BONUS: STM → Laughter

Differences between Mimicking and Non-Mimicking laughter in Child-Caregiver Conversation: A Distributional and Acoustic Analysis

Anonymous CogSci submission

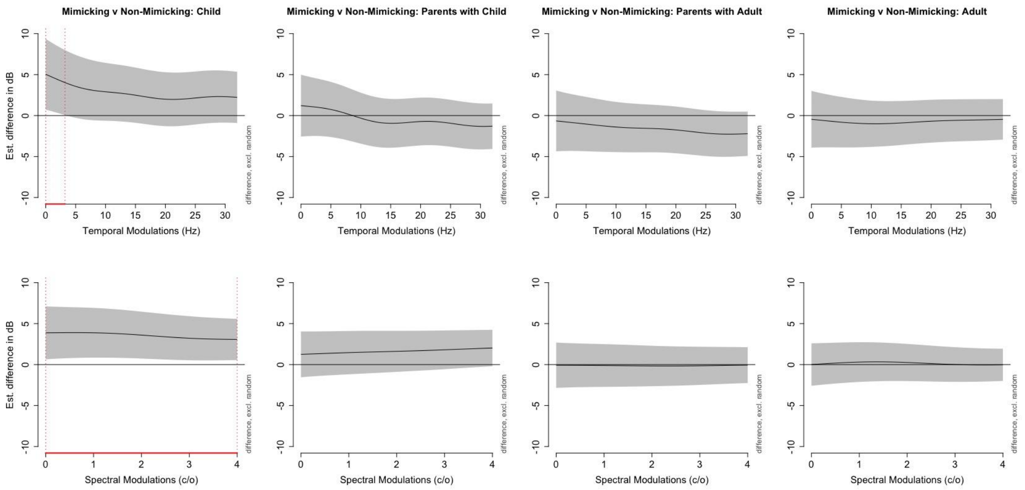
accepted to CogSci 2023

Question Acoustic differences between child/parent and parent/adult dyad when producing (non-)mimicking laughter?

STM Used to distinguish speech, laughter, and speech laughter (Ludusuan & Wagner, 2020)
GAMMs model variance over spectrum → opposed to fixed value

Results Children: M > N
M: Parents w/ children > adults

Reflections TM: 0-3 Hz → prosody → emotion
SM: 0-4 cyc/oct → arousal?



Bonus: Golf putting (1)


Experimental Brain Research (2020) 238:883–895
https://doi.org/10.1007/s00221-020-05757-3
RESEARCH ARTICLE
Online sonification for golf putting gesture: reduced variability of motor behaviour and perceptual judgement
Benjamin O'Brien[1] · Brett Juhas[1] · Marta Bieńkiewicz[1] · Frank Buloup[1] · Lionel Bringoux[1] · Christophe Bourdin[1]

Goal: Effects of online sonification of golf putting gesture on behavioral and perceptual performance with novices

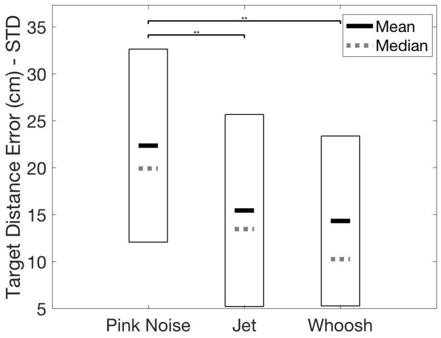
Method:

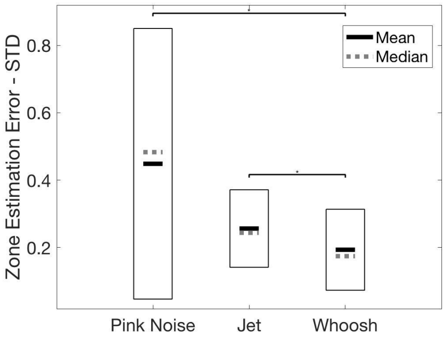
a. Develop 24 real-time sonifications based on club head speed:
   2 synthesizers * 2 modalities * 2 gains * 3 mappings

b. limit vision at impact

c. 5 sessions of 25 trials (random), target 3.5 m

	Brightness	Rhythmicity
Whoosh		
Jet		





✔ Relationship between performance & **movement variability?**

Bonus: Golf putting (2)

RESEARCH QUARTERLY FOR EXERCISE AND SPORT
https://doi.org/10.1080/02701367.2020.1726859
Routledge Taylor & Francis Group

Sonification of Golf Putting Gesture Reduces Swing Movement Variability in Novices

Benjamin O'Brien, Brett Juhas, Marta Bieńkiewicz, Frank Buloup, Lionel Bringoux, and Christophe Bourdin

Study Question: Effect of auditory guidance, feedback on performance & movement variability

Method:

a. Develop mean velocity profile (MVP), error

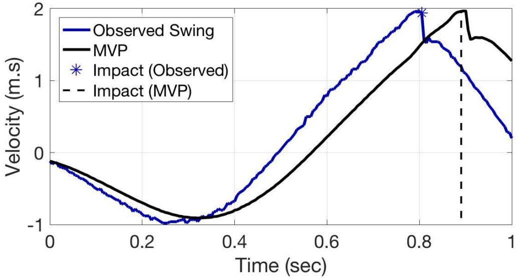
b. Develop auditory conditions

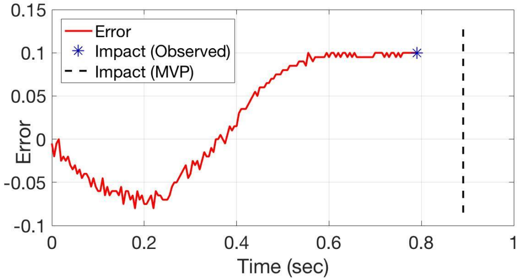
Control: Pink noise

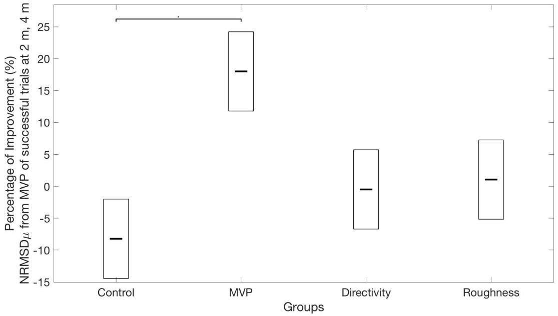
Auditory guidance
MVP → Frequency

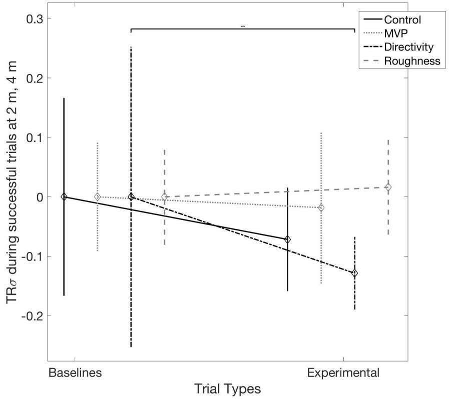
Auditory feedback
Error → Pan MVP signal

Auditory feedback
Error → Modulate MVP signal









Speech: Corpus

passage lu "La bise et le soleil" ou"La chèvre de M. Seguin" ´segmenté et étiqueté phonétiquement

Québec *singing corpus* – training corpus, Tremblay *et al*, 2018

N=146, ages between 20 & 89 years (mean 53.89 ± 19.82), 43% F, 49% singers

C2Si (head-neck cancer) - validation corpus, Woisard *et al.*, 2020

N=127
Patients (67%) : ages between 36 & 87 years (mean 60 ± 13), 45% F
Control: ages between 30 & 77 years (mean 65 ± 9), 61% F

AHN PARK (Dysarthric Parkisons) - validation corpus, Ghio *et al.*, 2018

N=316
Patients (65%) : ages between 29 & 85 years (mean 66.51 ± 9.63), 33% F
Control: ages between 38 &87 years (mean 62.68 ± 11.06), 49% F