

Head tracker using webcam for auralization

**Davi Rocha Carvalho, William D'Andrea Fonseca, Jacob Hollebon,
Paulo Henrique Mareze and Filippo Maria Fazi**



inter
noise 2021

① Introduction

Motivation

② The model

③ Results & Discussion

④ Conclusions

Outline

① Introduction

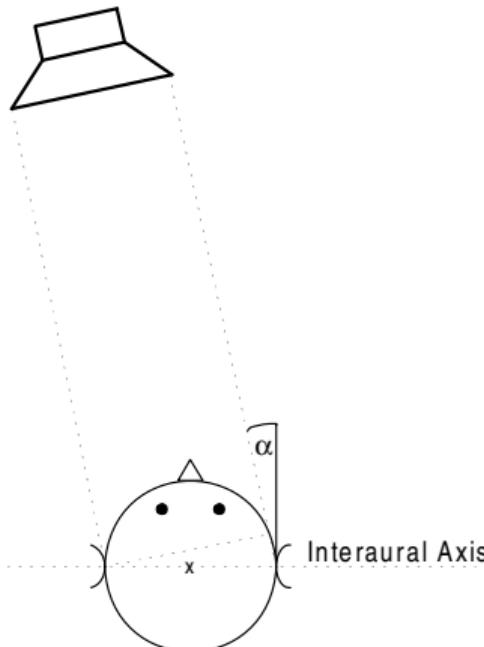
Motivation

② The model

③ Results & Discussion

④ Conclusions

Introduction: Binaural technology



Binaural Hearing:

- Natural listening;
- Consequence of having 2 ears;
- The **3D** acoustic perception of environments; and
- Allows sound sources **localization**.

Figure 1: Adapted from Blauert [1].

Binaural technology

Head-related Transfer Functions (HRTFs)

- Transfer function of the **pathway** between a **sound source** and the **listener's ear canal**.

Applications:

- 3D audio scenes on headphones and
- Virtual/augmented reality (VR/AR).

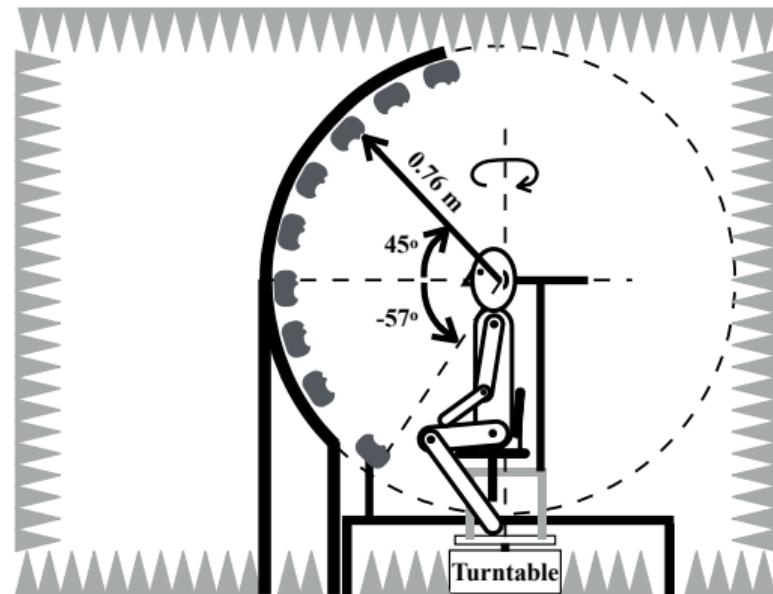


Figure 2: HRTF measurement setup [2].

Why use a head tracker?

Recording:

- Capture head movements during HRTF/BRIR measurements.

Reproducing:

- Real-world movement converted into virtual-world interaction.
 - Closer to natural perception than non-tracked auralizations;
 - Increased immersion to real-time rendered scenes; and
 - Correct source positioning in augmented reality (AR).

Head tracker in a binaural pipeline

- How to integrate a head tracker to a real-time auralization pipeline?

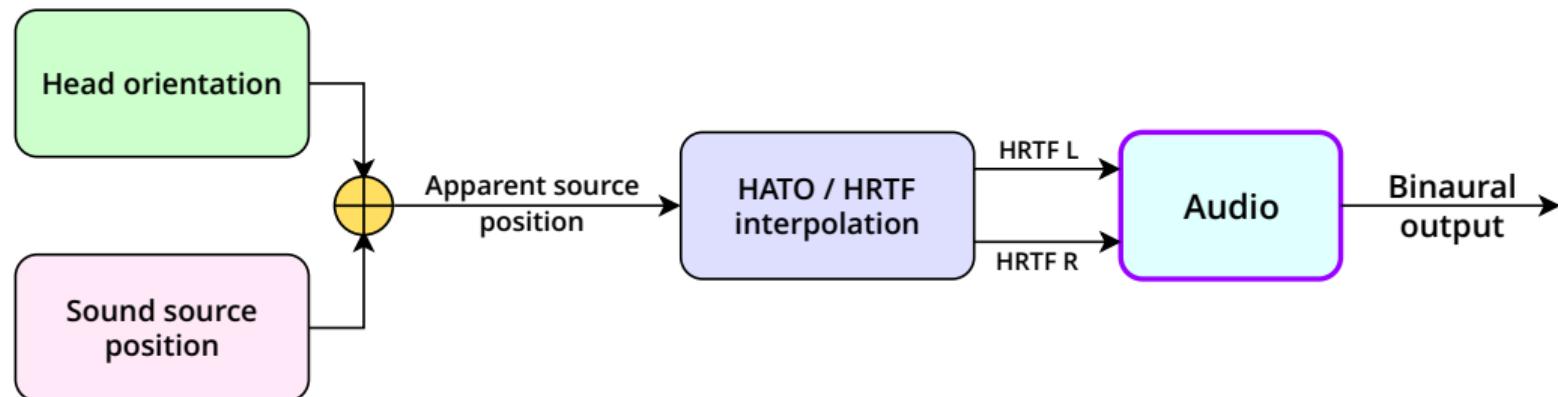


Figure 3: Simple auralization pipeline for binaural rendering with a head tracker.

Head tracker in a binaural pipeline

Hardware based

- ① The sensor is usually mounted to the listener's head;

Software based

- ① The sensor doesn't need to physically interact with the listener;

Head tracker in a binaural pipeline

Hardware based

- ① The sensor is usually mounted to the listener's head;
- ② Not free (material costs);

Software based

- ① The sensor doesn't need to physically interact with the listener;
- ② Open to use (without cost);

Head tracker in a binaural pipeline

Hardware based

- ① The sensor is usually mounted to the listener's head;
- ② Not free (material costs);
- ③ Limits depend on the hardware;
and

Software based

- ① The sensor doesn't need to physically interact with the listener;
- ② Open to use (without cost);
- ③ Limited to the camera opening angle;
and

Head tracker in a binaural pipeline

Hardware based

- ① The sensor is usually mounted to the listener's head;
- ② Not free (material costs);
- ③ Limits depend on the hardware; and
- ④ It can be damaged.

Software based

- ① The sensor doesn't need to physically interact with the listener;
- ② Open to use (without cost);
- ③ Limited to the camera opening angle; and
- ④ Without planned obsolescence.

Outline

① Introduction

Motivation

② The model

③ Results & Discussion

④ Conclusions

The model: Camera based approach

- Mediapipe's face mesh [3], a library for image AR (augmented reality):
 - Face detection and tracking via landmark identification.
- 468 landmarks.

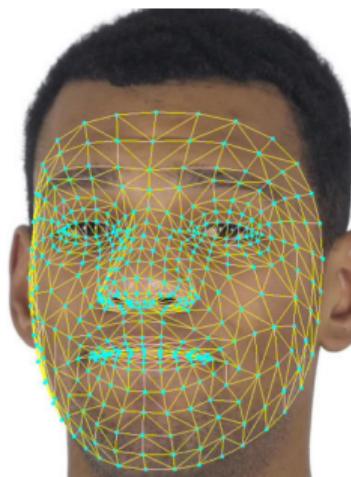


Figure 4: Face landmarks (adapted from [4]).

Camera based approach

- **34 out of the 468** landmarks are tracked
 - Face expressions influence and
 - Reduce processing overhead.

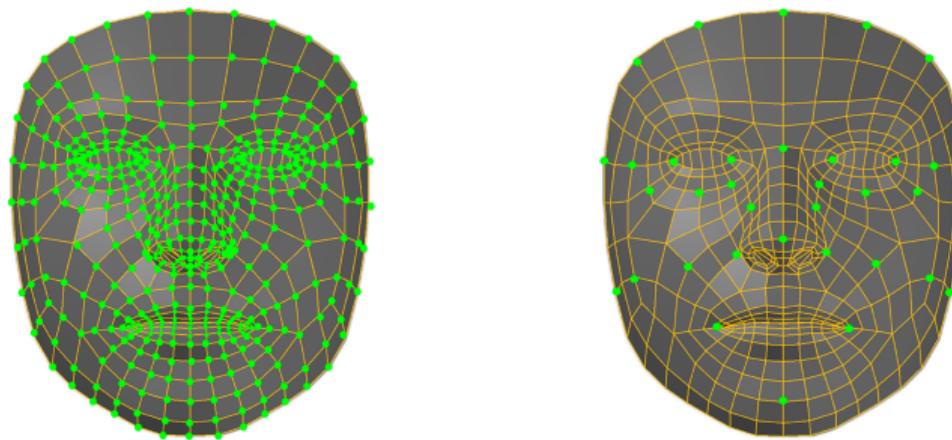


Figure 5: Comparison between 468 default landmarks and the 34 selected (adapted from [4]) inter noise 2021

Camera based approach

- Rotational angles extracted from landmarks are necessary for the auralization.

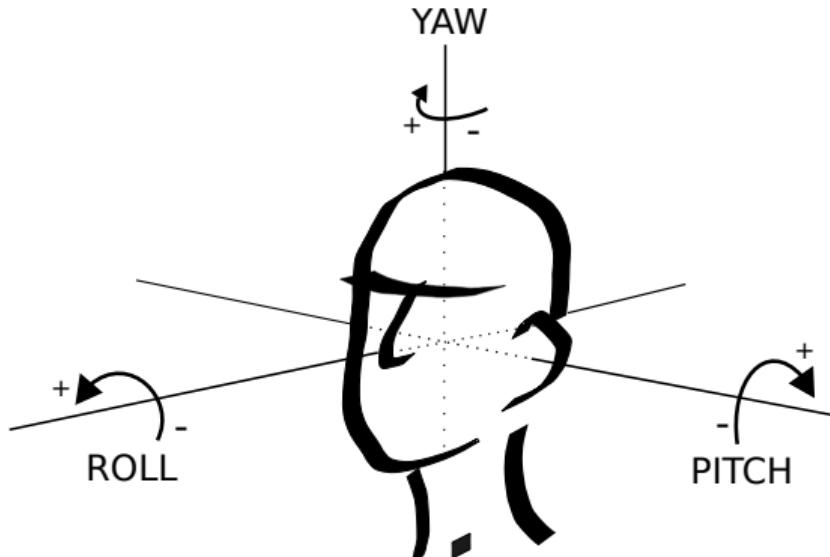


Figure 6: Yaw, pitch, and roll of the human head.

The model: Communication protocol

User datagram protocol (UDP)

- Continuous data stream;
- Low latency transmission; and
- Flexible message encoding.

Code 1: How to send data via UDP in Python.

```
import socket as skt
IP = '127.0.0.1'
PORT = 50050
s = skt.socket(skt.AF_INET, skt.SOCK_DGRAM)
coord = '{yaw},{pitch},{roll}'.format(yaw=-5,pitch=-2,roll=10)
s.sendto(coord.encode(), (IP,PORT))
```

Outline

① Introduction

Motivation

② The model

③ Results & Discussion

④ Conclusions

Results and Discussion: Hardware cross-validation



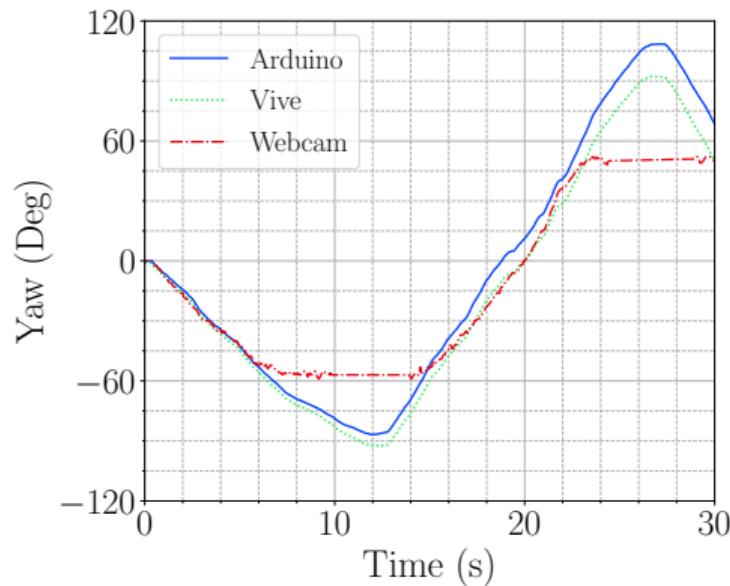
(a) Front view of the setup.



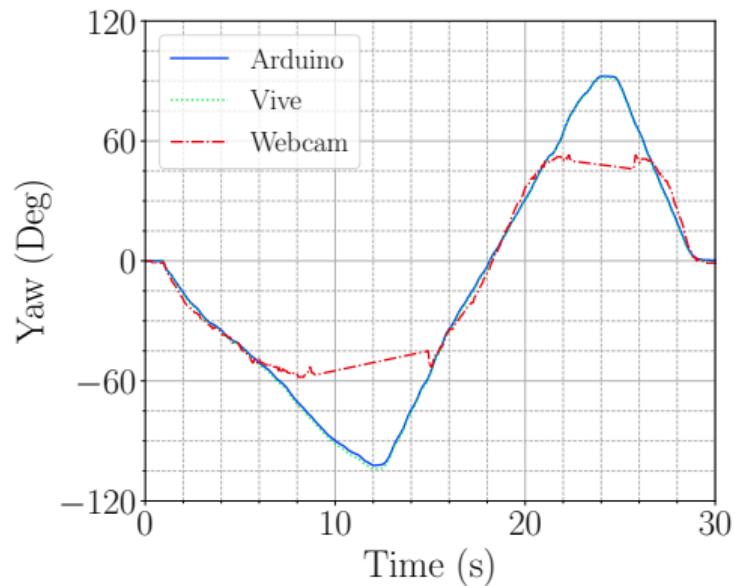
(b) Rear view of the setup.

Figure 7: Neumann KU-100 with the three head trackers — the Arduino, Vive, and webcam.

Hardware cross-validation



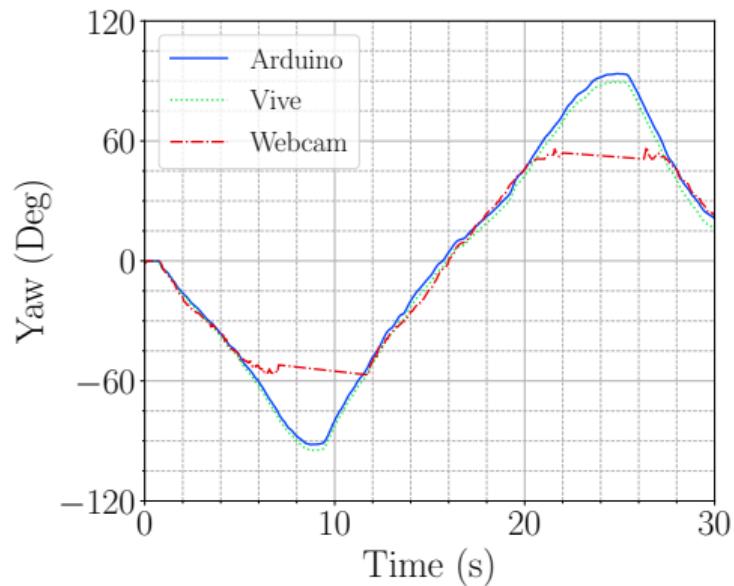
(a) Yaw Measurement 1.



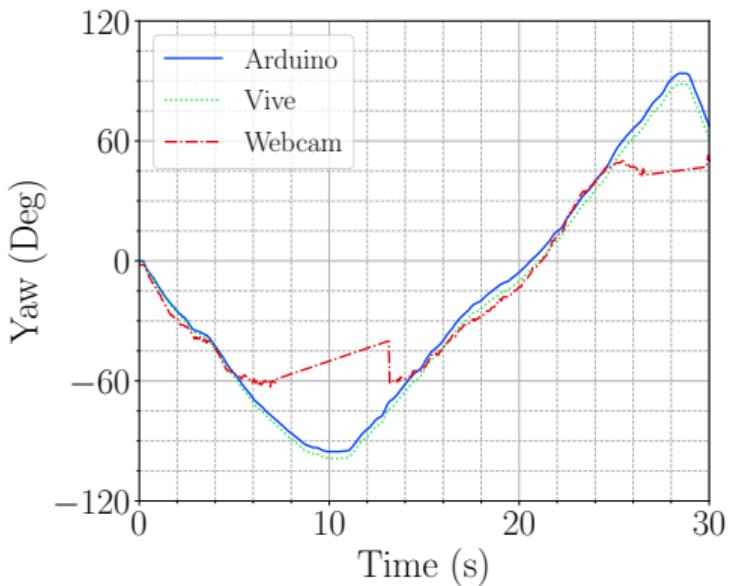
(b) Yaw Measurement 2.

Figure 8: Head tracker comparison for a yaw rotation, repeated measurements.

Hardware cross-validation



(a) Yaw Measurement 3.



(b) Yaw Measurement 4.

Figure 9: Head tracker comparison for a yaw rotation, repeated measurements.

Motion tracking amplitude

- Real subject motion extension

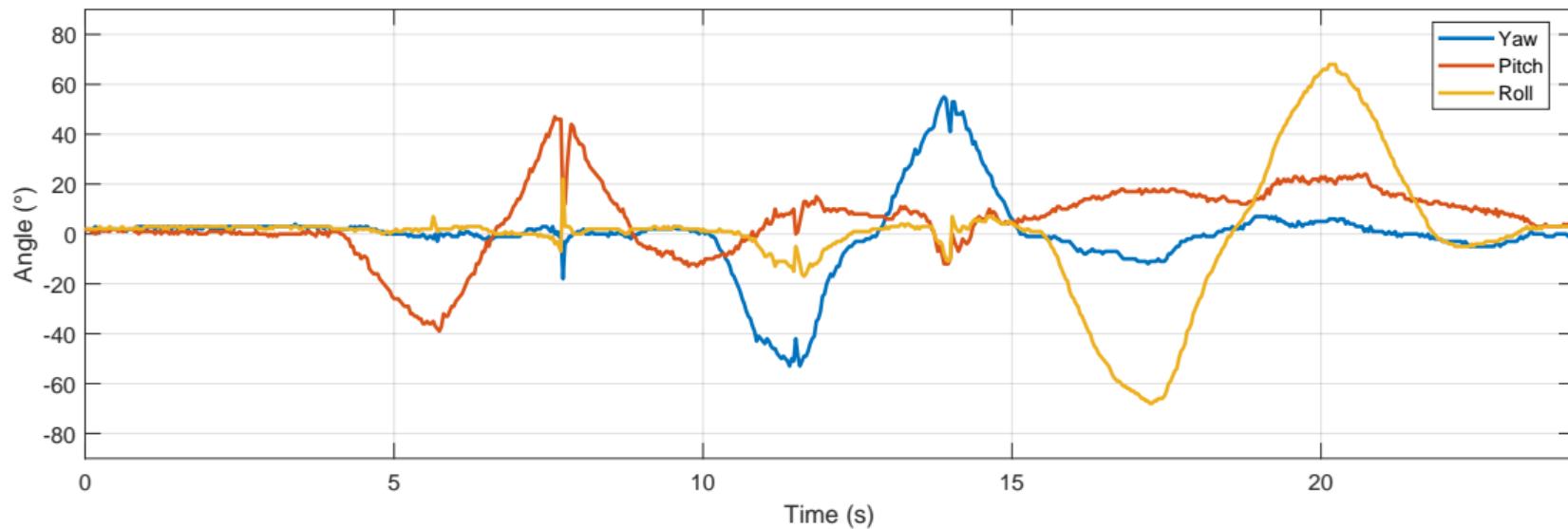


Figure 10: Head tracker motion extension measurement.

Results summary

Main characteristics:

- **Precision:** Good agreement with commercial hardware sensors;
- **Amplitude of motion tracking:**
 - Yaw $\approx 120^\circ$;
 - Pitch $\approx 80^\circ$; and
 - Roll: 360° .
- **Resolution:** Truncated to 1° (*tunable*);
- **Latency:** Yet to be determined.

Example: Real-time auralization

Videos

- <https://youtu.be/k5Od8YP294A>
- https://youtu.be/e_5j7NOdxMc

Outline

1 Introduction

Motivation

2 The model

3 Results & Discussion

4 Conclusions

Conclusions

- **Comfortable, convenient and affordable solution** than hardware implementations;
- **Comparison to commercial hardware** head trackers, demonstrate **good agreement**;
- **Limited range of coverage angles**: suitable to applications where the user needs to keep looking at the screen at all times; and
- Flexible communication system: change message output format according to necessity.
- Additional data of this project can be found in the GitHub repository
<https://github.com/eac-ufsm/internoise2021-headtracker> .

Future works

- Quantify system's latency;
- Explore the viability of expansion to a multi-camera system; and
- User friendly interface, including tweakable parameters.



Thank you!

Davi Rocha Carvalho, William D'Andrea Fonseca, Jacob Hollebon, Paulo Henrique Mareze & Filippo Maria Fazi.



davi.carvalho@eac.ufsm.br



EAC
ACOUSTICAL ENGINEERING | UFSM
www.eac.ufsm.br

UNIVERSITY OF
Southampton
isvr Institute of Sound
and Vibration Research



Amplifiers, Bolling Field, 1921.

References I

-  [1] Jens Blauert, ed. *Communication Acoustics*. Springer-Verlag Berlin Heidelberg, 2005. ISBN: 978-3540274377. DOI: [10.1007/b139075](https://doi.org/10.1007/b139075).
-  [2] Rahulram Sridharan, Joseph G. Tylka, and Edgar Choueiri. "A database of head-related transfer function and morphological measurements". In: *143rd AES Convention*. New York, NY, USA, Oct. 2017, pp. 1–5. URL: <https://www.aes.org/e-lib/browse.cfm?elib=19308>.
-  [3] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Ubweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. "MediaPipe: A Framework for Building Perception Pipelines". In: *CoRR abs/1906.08172* (2019). URL <http://arxiv.org/abs/1906.08172> or <https://github.com/google/mediapipe>. arXiv: [1906.08172](https://arxiv.org/abs/1906.08172).
-  [4] Yury Kartynnik, Artsiom Ablavatski, Ivan Grishchenko, and Matthias Grundmann. "Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs". In: *CoRR abs/1907.06724* (2019). arXiv: [1907.06724](https://arxiv.org/abs/1907.06724). URL: <http://arxiv.org/abs/1907.06724>.