

ILD and ITD Extraction with a Bio-Inspired Model Based on LSO and MSO

Lukas Driendl^{1,4}, Jörg Encke², Nelly von Puttkamer¹, Simon Schiele³, Tim Lüth³, Werner Hemmert¹

¹ *Bio-inspired Information Processing, Munich Institute of Biomedical Engineering, Technical University of Munich, 85748 Garching, Germany*

² *Medical Physics and Acoustics, Carl von Ossietzky University of Oldenburg, 26111 Oldenburg, Germany*

³ *Micro Technology and Medical Device Technology, Technical University of Munich, 85748 Garching, Germany*

⁴ *Email: lukas.driendl@tum.de*

Introduction

Most sound localization algorithms rely on correlation methods. They are inspired by a neuronal coincidence detection model with delay lines as postulated by Jeffress [1], which was actually found in barn owls later. However, recent research has shown that in mammals the decoding of interaural time differences (ITDs) in the medial superior olive (MSO) is based on excitation and contralateral inhibition [2, 3]. As shown in the PhD thesis of Encke, this neuronal circuit decodes the phase difference between left and right ear signals along the tonotopic axis provided by the cochlear filters [4].

In this paper we have developed and implemented a physical bio-inspired model of the detection of interaural level differences (ILDs) and ITDs. The system consists of a three-dimensional printed human head with pinnae. We fixed electret microphones in both ear canals from the inside of the head and recorded signals in a room and outdoors. We used a complex gammatone filterbank to extract both magnitude and phase of the ear signals. We calculated the magnitude difference (ILD in dB) between left and right ear to mimic processing in the left and right lateral superior olive (LSO) and the phase differences in the medial superior olive (MSO) to extract ITDs and evaluated the system.

Model Structure

Gammatone filters are commonly used in cochlear modeling to approximate the filtering behaviour of the peripheral auditory system. Measurements have shown that the cochlear impulse response resembled that of a gammatone function [5] in the time domain. However, it has to be noted that there are fundamental differences between gammatone filters and the filter characteristics of the hearing organ. Especially the low-frequency behaviour of gammatone filters deviate massively from the hearing organ: as seen in Figure 1, gammatone filters do not suppress very low frequencies whereas our hearing organ has at least two zeros at 0 Hz (one from the helicotrema and the second one from the velocity coupling of inner hair cell stereocilia). Unsufficient suppression of very low frequencies of gammatone filters could cause problems in real-life environments, for example, when a door is opened. In these cases, more realistic filter banks with appropriate low-frequency suppression should be chosen.

In this paper, we used a gammatone filterbank to decompose the incoming sound signal into 24 sub-bands,

spaced according to the equivalent rectangular bandwidth (ERB) scale. The ERB defines the bandwidth of each auditory filter as rectangular band-pass filters along the length of the basilar membrane and approximates human data at moderate sound levels as a linear function [6]:

$$\text{ERB}(f_c) = 24.7 \cdot (4.37f_c/1000 + 1) \quad (1)$$

Center frequencies f_c were chosen to span a range of 100 Hz to 4 kHz.

The gammatone impulse response is defined as the amplitude modulation of a cosine tone with the envelope of a Gamma distribution

$$g(t) = \begin{cases} at^{n-1}e^{-2\pi\text{ERB}t} \cos(2\pi f_c t + \phi) & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases} \quad (2)$$

Additional parameters of the gammatone function are:

- a : proportionality factor
- n : filter order
- ϕ : phase of carrier tone

The proportionality factor a was set such that each filter achieves a gain of 0 dB at its center frequency, visible in Figure 1, which shows the frequency response of the 24 implemented gammatone filters. The bandwidth of human auditory filter shapes is best approximated by a filter order of $n = 4$ [7, 8]. The starting phase ϕ of the carrier tone was left at 0° .

In order to retain the phase information of the signal, a complex notation is used instead of a cosine wave, resulting in a complex impulse response:

$$g(t) = \begin{cases} at^{n-1}e^{-2\pi\text{ERB}t} e^{j2\pi f_c t + \phi} & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases} \quad (3)$$

All model calculations were performed in MATLABTM [9] using the Auditory Modeling Toolbox [10].

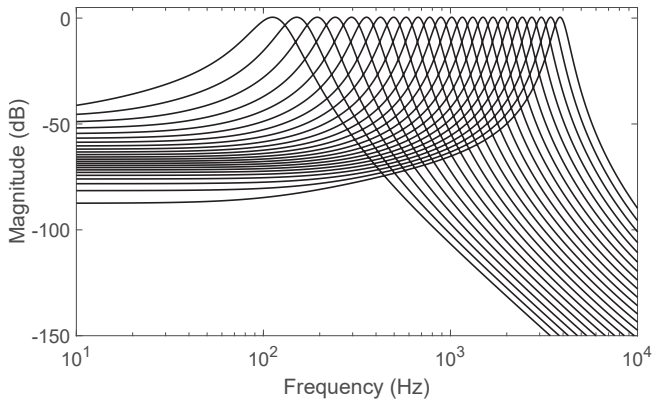


Figure 1: Frequency response of complex gammatone filterbank with 24 filters spanning 100 Hz to 4 kHz.

The input signals for left and right ear are filtered by a left and a right filterbank, leading to channel outputs g_l and g_r for a given center frequency f_c of

$$g_l(t) = a_l(t) \cdot e^{j\phi_l(t)} \quad (4)$$

$$g_r(t) = a_r(t) \cdot e^{j\phi_r(t)} \quad (5)$$

with instantaneous phases ϕ_l and ϕ_r , magnitudes a_l and a_r .

ITD Extraction

Stimulus signals reaching the head from an azimuth angle $\alpha \neq 0^\circ$ or $\alpha \neq 360^\circ$ will result in waveforms reaching the ears with a small phase shift between them, called interaural phase delay (IPD). As human ITD sensitivity is frequency limited [11], we only used filter channels with center frequencies up to 1.016 kHz (channels 1 to 13) in order to extract IPDs. As the filterbank produces complex outputs, the IPD can directly be extracted by taking the difference in phase between left and right signal.

$$\text{IPD}(t) = \arg \left(\frac{g_l(t)}{g_r(t)} \right) = \phi_l(t) - \phi_r(t) \quad (6)$$

For filters above 600 Hz (channel 10 and above) the phase difference can be larger than $\pm 180^\circ$ due to the dimensions of the physical head model and the speed of sound. To still utilize higher frequency channels for IPD and ITD detection, IPDs for the affected channels were unwrapped based on information obtained from low frequency channels where the phase is unambiguous.

Phase delays are then converted into ITDs by their respective filter center frequencies f_c , resulting in

$$\text{ITD}(t) \approx \frac{\text{IPD}(t)}{2\pi f_c} \quad (7)$$

Eq. 7 is an approximation as each filter has a given filter bandwidth determined by Eq. 1, therefore leading to errors towards the filter borders.

For localization tasks, ITDs are commonly mapped to specific azimuth angles based on simple look up tables. In order to simplify the gammatone model further for such tasks, the 13 individual ITD output channels were combined using a weighted sum. The weighting was performed according to the signal power as well as ITD variance of each model output, ensuring that sections with high signal power as well as channels with less ITD variance are weighted strongly.

Azimuth measurement angles are defined here such that positive angles correspond to sound sources located to the left of the head model and negative angles to sound sources located to the right. Due to Eq. 6, negative IPD and ITD values therefore occur for sound waves coming from the left and vice versa.

ILD Extraction

Level differences in the model were calculated as the difference in instantaneous magnitude between left and right ear channel as

$$\text{ILD}(t) = 20 \cdot \log_{10} \left[\frac{a_r(t)}{a_l(t)} \right] \text{ dB} \quad (8)$$

where negative values correspond to sound sources located to the left of the head model and positive values for right sound source locations.

Physical Head Model

First model tests were performed using manually delayed speech samples as well as audio samples convolved with head related transfer functions (HRTFs) measured using a KEMAR dummy head from the MIT Media Lab [12]. More realistic tests were performed using a 3D printed physical head model with electret microphones fixed at the outer edge of the ear canals (Figure 2). The head model was printed using selective laser sintering (SLS) with PA2200 on a EOS Formiga P100 machine, a powder based additive manufacturing process.

The signals recorded by the two microphones were taken as the binaural input for the gammatone filterbank, resulting in left and right output channels.



Figure 2: 3D printed head model used for freefield and room measurements. Two electret microphones have been inserted at the entrance of the left and right ear canal. The head is life-sized and modelled after a CT-scan. Dimensions are 20 cm x 18 cm x 16 cm (length x width x height). Fabricated at the Institute of Micro Technology and Medical Device Technology, Technical University of Munich.

Measurements were recorded in a room and in an open space outside (freefield), as shown in Figure 3. Both setups consisted of a speaker mounted 150 cm away from the head model, with both positioned at a height of 100 cm above ground. The freefield measurement was conducted in order to reduce the impact reflections have on the model output. As no precedence effect was incorporated at this time, the model output for ITDs and ILDs was highly sensitive to reflections, resulting in large output swings. This was highly reduced for the freefield measurement.



Figure 3: Measurement setup in freefield conditions to reduce impact of reflections on model output. Speaker was oriented towards open space (lake) with a distance of 150 cm to the physical head model.

Results and Evaluation

Model ITD outputs of the previously described weighted sum are shown for multiple freefield measurement angles in Figure 4. Notably, ITD outputs between the chosen measurement angles were nicely separated, allowing for clear sound source localization. Means and standard deviations of the same measurement ITD outputs are shown in Figure 5.

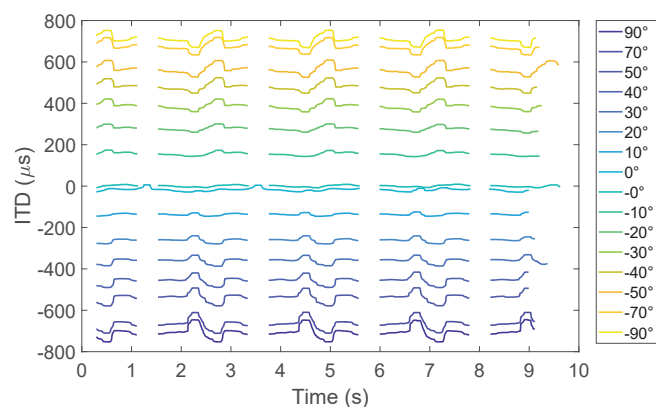


Figure 4: Weighted sum ITD output of gammatone model for measurement angles between 90° and -90° (from bottom to top) for a test sentence taken from the OLSA speech test. Measurement performed in freefield conditions.

Larger angles resulted in increased standard deviations of the ITD output, also visible in Figure 4, as output swings increased with the azimuth angle.

Standard deviations for ITDs of individual filter channels are shown in Figure 6, where (a) represents a model worst case at a measurement angle of 90° . The standard deviation is highly dependent of channel center frequency. The measurement at 0° shown in (b) has overall much smaller standard deviations (note different scale in Figure 6a and b).

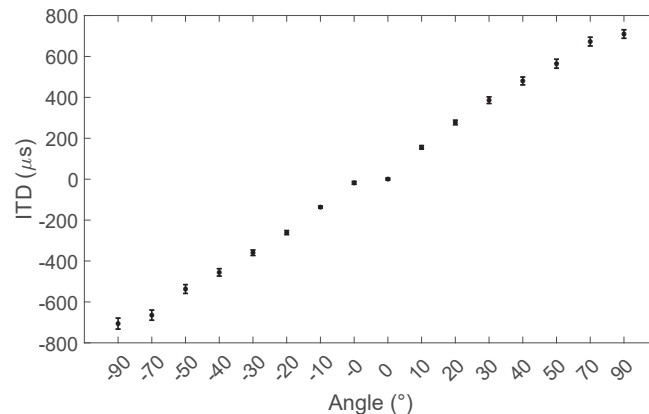


Figure 5: Means and standard deviations of weighted sum ITD model output for measurement angles between -90° and 90° .

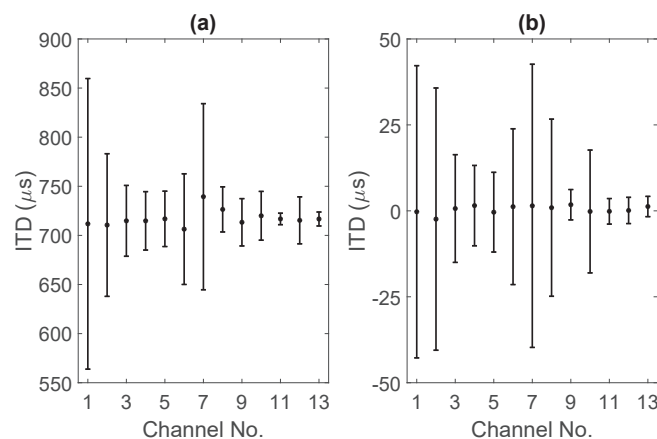


Figure 6: Means and standard deviations of individual ITD channels for (a) measurement angle of -90° and (b) measurement angle of 0° .

ILDs were calculated for the same freefield measurement as for the ITD evaluation with azimuth angles between -90° and 90° . Figure 7 shows the ILD trend of each channel over multiple measurement angles as mean values. Level differences increased with azimuth angle, however at measurement angles close to $\pm 90^\circ$, ILD curves started to overlap and as such, mapping of sound source localization would be difficult here. ILDs generally were not completely symmetric because the head model and microphone placement was also not perfect.

Freefield and room measurements mostly differed in the individual ITD channel outputs, as reflections present in the room measurements produced large IPD swings in some channels and therefore ITD output swings. The weighted sum, however, averaged out a large part of these

output distortions, leading to comparable combined ITD outputs between freefield and room measurements for all tested azimuth angles.

Lastly, as expected, manually delayed speech signals as well as audio signals convolved with HRTFs resulted in much cleaner ITD and ILD model outputs, mostly due to the absence of reflections as the impulse responses were measured in an anechoic chamber.

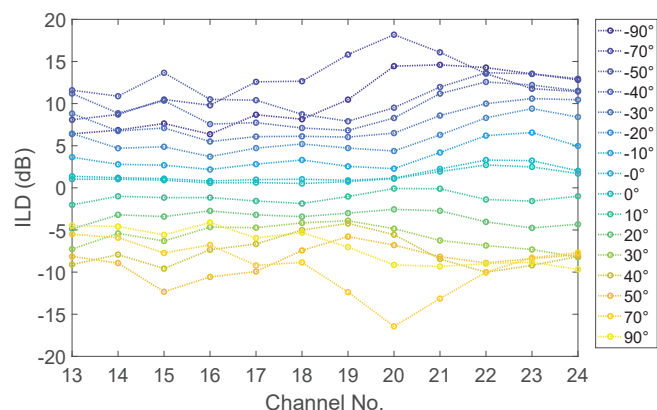


Figure 7: Means of ILD channel outputs for measurement angles between 90° and -90° (from bottom to top).

Conclusions

A simplified model of the MSO and LSO for sound source localization, based on a 24-channel complex gammatone filterbank has been proposed in this paper. The model can efficiently extract interaural cues including IPDs, ITDs and ILDs. ITDs of 13 low frequency filter channels were combined using a weighted sum algorithm resulting in a stable filterbank output which proved resistant against reflections obtained in regular room measurements. We believe computationally efficient models such as the complex gammatone filterbank model can prove valuable for sound source localization tasks for example in robot hearing.

Acknowledgements

This study was funded in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number: 415658392.

References

- [1] Jeffress, L. A.: A place theory of sound localization. *Journal of comparative and physiological psychology* (1948), 35-9
- [2] Magezi, D. A., & Krumbholz, K: Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *Journal of Neurophysiology* (2010), 1997-2007
- [3] Grothe, B., Pecka, M., and McAlpine, D.: Mechanisms of sound localization in mammals. *Physiological Reviews* (2010), 983-1012
- [4] Encke, J. & Hemmert, W.: Extraction of Inter-Aural Time Differences Using a Spiking Neuron Network Model of the Medial Superior Olive. *Frontiers in Neuroscience* (2018), 140

- [5] Johannesma, P.I.M.: The pre-response stimulus ensemble of neuron in the cochlear nucleus. *Symposium of Hearing Theory* (1972), 58-69
- [6] Glasberg, B. R., & Moore, B. C. J.: Derivation of auditory filter shapes from notched-noise data. *Hearing Research* 47 (1990), 103-138
- [7] Patterson, R., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., & Allerhand, M.: *Complex Sounds and Auditory Images*. International Symposium on Hearing (1992), 429-446
- [8] Patterson, R.D. and Moore, B.C.J.: Auditory filters and excitation patterns as representations of frequency resolution. *Frequency Selectivity in Hearing* (1986), 123-177
- [9] Matlab v9.6.0 (R2019a), URL: <https://de.mathworks.com/>
- [10] Auditory Modeling Toolbox 1.0, URL: <https://amtoolbox.org/>
- [11] Brughera, A., Dunai, L. & Hartmann, W. M.: Human interaural time difference thresholds for sine tones: The high-frequency limit. *The Journal of the Acoustical Society of America* (2013), 2839-2855
- [12] HRTF Measurements of KEMAR Head (MIT Media Lab), URL: <https://sound.media.mit.edu/resources/KEMAR.html>