EENG580 Activity 5: Audio Position Tracking

Activity summary

Overview: Design a system to track the position of a speaker using microphones

Setting: Develop in class, then refine after class (any location).

Curricular elements: both tinkering and gaming

Prerequisites: basic familiarity with MATLAB

Topics/concepts covered: Fourier properties, sampling, upsampling, downsampling

Learning outcomes: After completing this activity, students should be able to:

- Use the time shift Fourier property to efficiently compute time delays
- Downsample, upsample, and interpolate a signal

Expected time to complete: two 2-hour lab sessions and 4-8 hours of work outside of class

Required hardware/materials: Minimum: two microphones attached to a computer in such a way that MATLAB can query them simultaneously.

Required instructor interaction: partially supervised, with occasional guidance

Common mistakes/pitfalls: If all students are using different hardware it can be difficult for the instructor to help debug problems. The instructor should verify correct operation of the devices before giving them to students. If possible, the instructor should test the audio recording capability on both PCs and Macs.

Method of assessment: instructor graded, based on final product

Contributors: Richard Martin, Andrew Klein, and Cody Watson

 This work was supported by ONR grant N000141512442, *Contact: andy.klein@wwu.edu and richard.martin@afit.edu.*

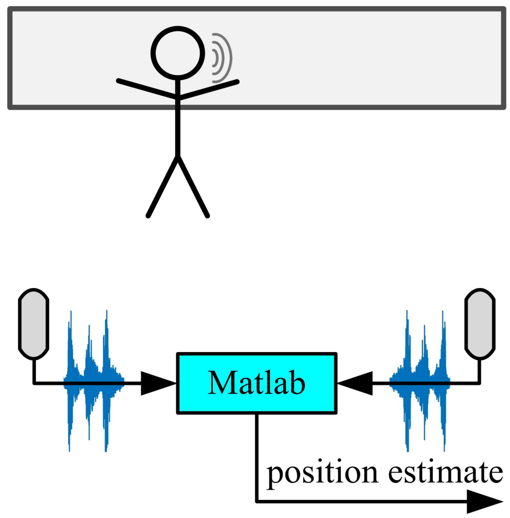

Figure 1: Geometry of an audio position tracking system.

1 Introduction

This activity will serve as a capstone for the course. You will use similar shift estimation methods as in "Activity 2: Image Registration," and you will build on your audio processing expertise from "Activity 1: Acoustic Impulse Responses," "Activity 3: Audio Filtering," and "Activity 4: Acoustic Wireless Communication System." Your goal is to use two microphones to track the position of a lecturer in front of a white board in real time, as shown in Fig. 1.

The global positioning system (GPS) operates using similar measurements. In GPS, the receiver uses the time of flight from multiple satellites to the receiver to estimate a 3D position and a clock offset. In your case, you will be estimating the position of a transmitter using multiple receivers rather than the other way around, but the idea is the same. To simplify the problem and require fewer measurements, you will reduce the problem to 1D by assuming that the lecturer is located somewhere on a line parallel to and 1 m away from the whiteboard. Ideally, the microphones will be placed at the same elevation as the audio source. You will also need to know the exact positions of the two microphones relative to the whiteboard.

2 Getting started

Before this activity begins, you are expected to have studied the following material in advance: sampling, downsampling, upsampling, and interpolation. Videos covering these concepts are available on the course website, and you are encouraged to read additional background material in any of the suggested reference textbooks.

Pick a lab partner for this activity. You will collaborate on the process, but you will write up individual reports and receive separate grades. Obtain two microphones and two audio jack to USB adapters from the instructors. The microphones are the same ones you used in the previous audio activities. Unless noted otherwise, the basic audio equipment used in this class will be comparable to the following:

- Microphones: Cyber Acoustics Desktop Unidirectional Microphone (CVL-1064), about \$7 each
- Adapters: Sabrent USB External Stereo Sound Adapter for Windows and Mac, about \$7 each

Ideally, everyone will use the same type of equipment rather than audio gear embedded in laptops, in order to maintain a level playing field. Plug in your equipment, make sure each device is turned on, and make sure the recording volumes are both set as high as possible.

Explore the following MATLAB commands:

- `audiodevinfo` (new!)
- `audiorecorder`
- `record`
- `stop`
- `getaudiodata`

The explanation in the help files should be straightforward; the only parameter you will need to choose is the sampling rate `fs`. Use your knowledge of Nyquist sampling to pick a good sampling rate for human speech.

Progress to the point where you can record a snippet of speech from both microphones simultaneously, plot both recordings on subplots within the same figure, and verify that they are similar except for a small time delay. If you can accomplish these tasks, you have sufficient understanding of the hardware interface to complete the entire activity.

3 Main tasks

Consider a measurement in which you determine that the distance from the speaker to microphone 1 is 2 m meters less than the distance from the speaker to microphone 2. If microphone 1 is located at $(0m, 0m)$, microphone 2 is located at $(3m, 0m)$, and the speaker is somewhere on the line $y = 5m$, where is the speaker? Once you figure that out, solve the problem more generally for arbitrary positions and time delays. If you couldn't assume that the speaker was on a line next to the whiteboard, what is the locus of points where the speaker could be? (Hint: it is a conic section.)

Have the speaker stand equidistant from the two microphones and record on the both simultaneously. Using techniques from "Activity 2: Image Registration," compute the time delay between the two recordings. There may be some non-zero delay due to the fact that matlab starts one recording slightly before the other. Experiment with this to calibrate your system; that is, figure out what the excess delay is so that you can subtract it out from future measurements.

Once your system is calibrated, take data in a variety of configurations. Use the geometrical modeling to map each calculated delay to a position estimate. Does it matter where you place the speakers? How accurate is your position estimator? Does it provide equal accuracy when the speaker is at the middle and at the endpoints of the line?

Now use sampling rates of 50 kHz for one microphone and 63 kHz for the other. You will have to resample one of the digital sequences to get them both to the same effective sampling rate before you can compute time delays. (This mimics real-life scenarios where you might be able to only buy samplers with fixed rates that are not the rates that you want.) The simplest approach is to upsample once, interpolate, and downsample once, but that involves a heavy amount of computation. A better approach is to note that

$$\frac{63}{50} = \frac{3 \cdot 3 \cdot 7}{2 \cdot 5 \cdot 5}, \tag{1}$$

so you can interlace three upsamplers and three downsamplers. If you do it right, you should never have to increase your storage use to more than seven times what you started with for the recording that you are resampling. Just keep in mind that if at any point in the chain you have downsampled by a larger amount than you have upsampled, then you will induce aliasing. Also, as you work through the resamplign and filtering, make sure that you do not inadvertently shift the data. To make sure, you can repeat the initial step of making a recording with the speaker equidistant from the two microphones.

Once you have sorted out the resampling process (either with two up/downsmaplers or six), re-validate your position tracker that now operates with two different sampling rates.

4 Deliverables

You will turn in a report written using the LaTeX word processing system. A template is available on the course webpage. The report should conform to the style guide for IEEE Signal Processing Society conferences, such as ICASSP, ICIP, or GlobalSIP. At a minimum, explain your measurement methodology, your data processing methodology, and your results. Both format and content matter – in particular, use correct grammar and spelling, and revise your text as needed to make sure you are explaining yourself coherently. Include code sparingly, if at all; in general, it is better to include an algorithm in pseudo-code as a figure.

You should strive to always explain why you did what you did, not just what you did. Your results should all be repeatable – maybe not exactly, but comparably. That is, another researcher may not have access to the exact same environment as you or record the exact same realization of random noise, but you want to enable them to follow the exact same process as you. Analysis is also highly encouraged – explain why the results are what they are, not just what they are. For this activity, photos or sketches of the geometry may help with that.

To improve your grade, include a more detailed and coherent explanation of your results, try a more clever or innovative approach, and include more data and analysis. Feel free to diverge somewhat from the stated tasks if you think the situation warrants it. “Out of the box” thinking will always be rewarded.

5 Competitions

The written report constitutes 90% of the grade, and the remaining 10% will be based on competitions. Competitions will include the following:

- Minimum position error: devise a method to measure “truth” data and to evaluate the root mean squared error (RMSE) of your position estimates. The class will vote to determine which evaluation methodology best convinces them of having the lowest RMSE. To be eligible, your system must use two different sampling rates rather than the initial design with matched sampling rates.
- Implement your tracking in real time. The class will vote to determine the best system design, in terms of its accuracy, how polished the final product is, and how well the position estimates are combined with some other visual display.
- Implement some other extension, such as using a third microphone to improve performance or compute a 2D position estimate, track two speakers at once, use a tracking filter (e.g. the Kalman filter) to actually create a position track rather than simply a sequence of estimates, or any other cool extension you can think of. Again, the winner will be determined by class vote.

The first place group in each category will receive a 10% bonus (for each group member) and the second place group members will each receive 5%; and no individual may receive more than 10% total. Based on typical class sizes, that means about a quarter of you will get 10% and another quarter will get 5%.