2022

Poster: Acoustic Source Localization Using Straight Line Approximations

Swarnadeep Bagchi
Technological University Dublin, d18128352@mytudublin.ie

Ruairí de Fréin
Technological University Dublin, ruairi.defrein@tudublin.ie

Funder: SFI

Acoustic Source Localization Using Straight Line Approximations

Swarnadeep Bagchi, Ruairí de Fréin

Technological University Dublin

Abstract

- Investigation of a delay estimation technique called tiled-Elevatogram to estimate delays of two dominant sources.
- We have tried to refine this technique using image procession methods in order to locate the delay-kinesics and called it chain-coded Elevatogram.
- The chain-coded Elevatogram technique performed poorly when compared to its predecessor.

Problem Statement

Can the tiled-Elevatogram [1] technique estimate the delays of two dominant sources in an anechoic mixing scenario ?

Previous Work

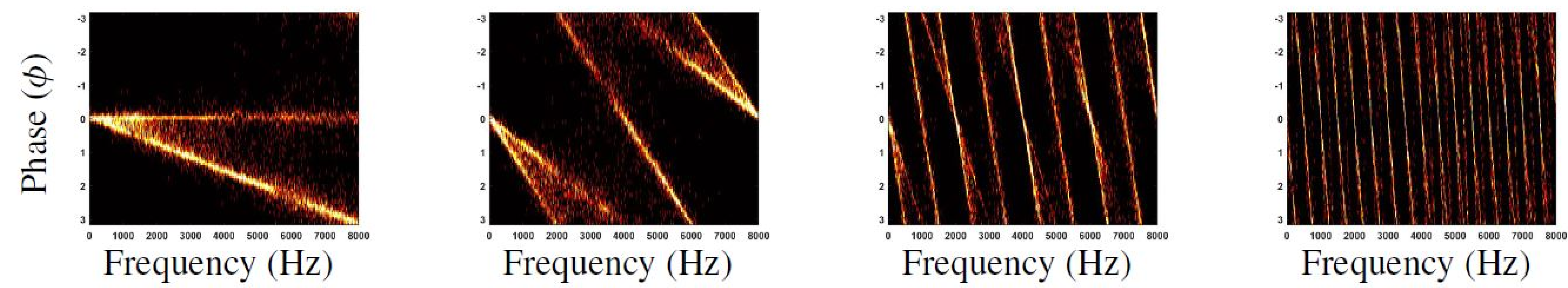
The tiled-Elevatogram [1] relied on the proximity of a particular microphone to the speech of interest.

Our Contribution

- Investigating and establishing the fact that the tiled-Elevatogram is well suited when two speech sources are equally dominant on that particular microphone.
- Based upon the original method, we have developed another method called chain-coded Elevatogram [2] by applying image processing techniques.
- The chain-coded Elevatogram method analyzes the delay patterns of the bright lines in Figure 1 as it meets the frequency axis.

Methodology & Experiments

Real speech utterances from the TIMIT database of $F_s = 16$ kHz, [3] are used. A K-sample FFT Hamming window was used where $K = 2048$ and $L = 100$.



(a) $\delta_1 = 0, \delta_2 = 1$ (b) $\delta_1 = 2, \delta_2 = 4$ (c) $\delta_1 = 16, \delta_2 = 8$ (d) $\delta_1 = 64, \delta_2 = 32$

Figure 1:The delays, δ_j are in samples. The slanting lines are observed for real speech utterances, where each line corresponds to a source, $s_j[n]$. The larger the delay, the more often the lines get phase wrapped. They indicate the energy concentration of a particular phase bin as a function of frequency.

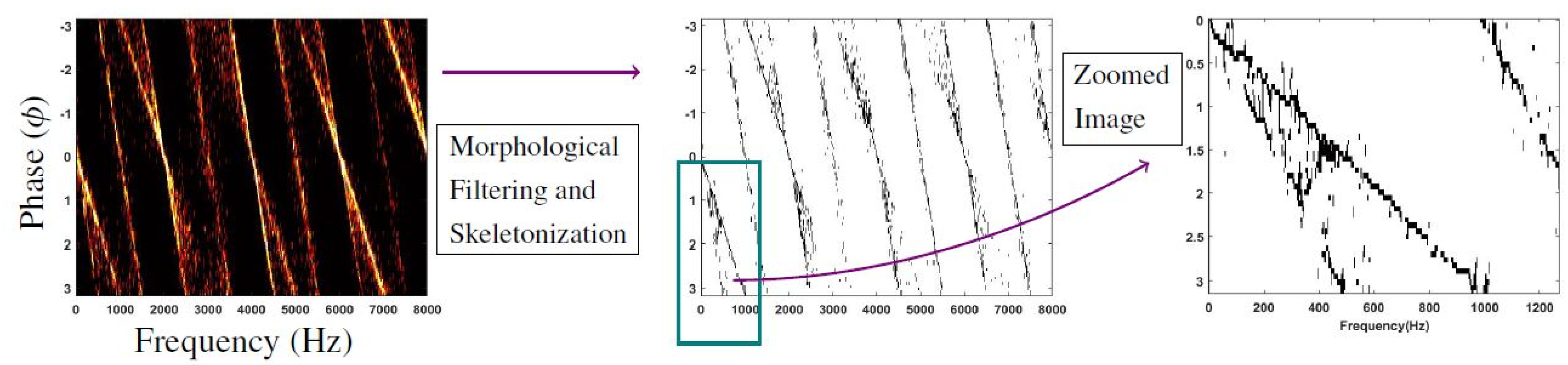


Figure 2:Morphological filtering and skeletonization. The two sources get phase wrapped first at locations 1000 Hz and 500 Hz. These lines are known as delay-kinesics.

Two lines corresponding to two sources start from the origin and bifurcates. We calculate the slope, m, pertaining to the two sources. The subsequent parallel line starts where the previous line ends. The coordinates of this line is computed using the formula: $m = \frac{y_2 - y_1}{x_2 - x_1}$. The only unknown is y_2. We notice that $x_1 = 1$ and $x_2 = L$. The subsequent parallel lines are calculated in the same procedure.

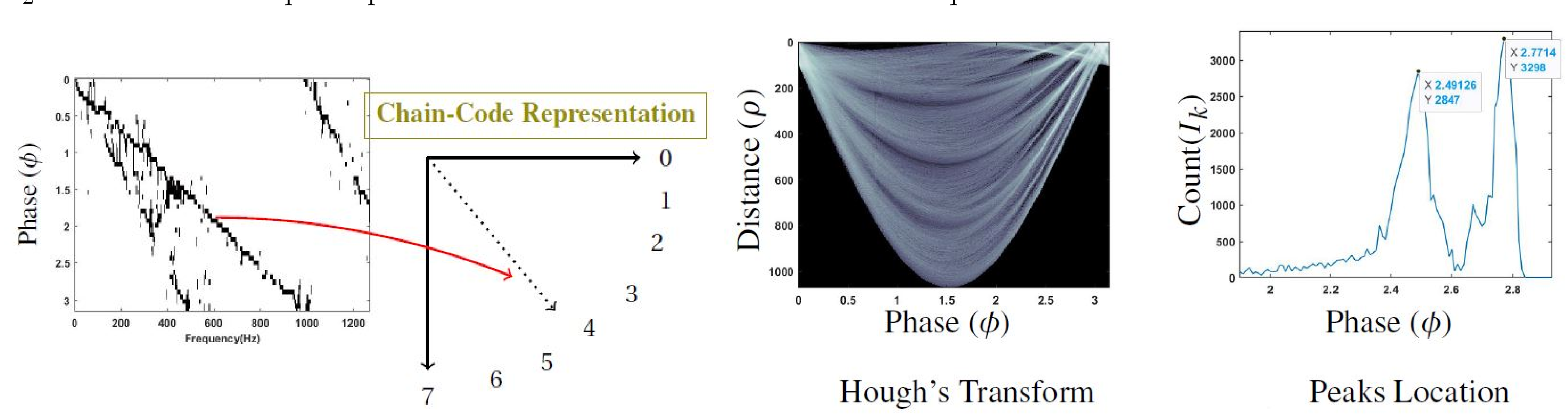


Figure 3:Two figures on LHS: Each code in $0 - 7$ corresponds to a particular direction. RHS figure most two most prominent peaks are observed at $\phi = 2.49$ and 2.77 rads. These corresponds to two accumulator cells getting two highest votes.

We skeletionize the phase-frequency matrix, $\mathbf{P} \in \mathbb{R}^{L \times K}$, without loss of significant information. Delay is estimated using the formula: $\hat{\delta} = -\frac{K}{L} \tan(\phi)$.

Results

Assuming ground truth are $\delta_1 = 16$ and $\delta_2 = 8$ in samples, the prominent peaks in Figure. 3 are at $\phi = 2.49$ and $\phi = 2.77$ rads. Now, substituting these values in $\hat{\delta}$ formula, we get $\hat{\delta}_1 = 15.6$ samples and $\hat{\delta}_2 = 7.94$ samples. The mean absolute differences between actual δ and estimated $\hat{\delta}$ are 0.4 and 0.06 samples, respectively.

Conclusion

- The limitation of the chain-coded Elevatogram is that it quantizes the direction of search for the continuity of a bright straight-line on the phase-frequency matrix. This results in a limited resolution for delay estimation. Unlike the tiled-Elevatogram, our chain-code approach performs unsatisfactorily for high delay estimates as the lines get so steep that it is impossible to differentiate the sources.
- Potential application of these techniques are in source separation and the hearing aid industry.

References

[1] Ruairí de Fréin.
Tiled time delay estimation in mobile cloud computing environments.
In *IEEE ISSPIT*, pages 282–287, 2017.

[2] Bagchi Swarnadeep, de Fréin Ruairí.
Acoustic Source Localization Using Straight Line Approximations.
In *IMVIP22*, 2022.

[3] John S Garofolo.
TIMIT acoustic phonetic continuous speech corpus.
Linguistic Data Consortium, 1993, 1993.

Acknowledgements

This paper has emanated from research supported in part by a Grant from Science Foundation Ireland under Grant number 18/CRT/6222 and Grant Number 15/SIRG/3459.