2022-09-21

Evaluating large delay estimation techniques for assisted living environments

Swarnadeep Bagchi
Technological University Dublin, d18128352@mytudublin.ie

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

Funder: SFI

Evaluating large delay estimation techniques for assisted living environments

Bagchi, Swarnadeep and de Fréin, Ruairí
Ollscoil Teicneolaíochta Baile Átha Cliath,
Campas na Cathrach,
Ireland.
web: https://robustandscalable.wordpress.com

in: Electronics Letters. See also BibTeX entry below.

BibTeX:

```
@article{deFrein22Evaluating,
author = {Bagchi, Swarnadeep and de Fr\'{e}in, Ruair\'{i}},
title = {Evaluating large delay estimation techniques for
assisted living environments},
journal = {Electronics Letters},
volume = {58},
number = {22},
pages = {846-849},
doi = {https://doi.org/10.1049/ell2.12624},
url = {https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ell2.12624},
eprint = {https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/ell2.12624},
year = {2022}
}
```


document created on: January 13, 2023
created from file: sample.tex
cover page automatically created with CoverPage.sty
(available at your favourite CTAN mirror)

Evaluating Large Delay Estimation Techniques for Assisted Living Environments

Swarnadeep Bagchi and Ruairí de Fréin

Phase wraparound due to large inter-sensor spacings in multi-channel demixing limits the range of relative delays that many time-frequency relative delay estimators can estimate. We evaluate the performance of a large relative delay estimation method, called the Elevatogram, in the presence of significant phase wraparound. This paper compares the Elevatogram with the popular relative delay estimator used in DUET and the brute-force approach in D-AdRess and analyzes its computational efficiency. The Elevatogram can accurately estimate relative delays of speech signals of up to 800 samples, whereas DUET and D-AdRess were limited to delays of 7 and 35 samples, given a sampling rate of 16 kHz. Monte Carlo trials on 1000 real speech utterances show that the Matlab execution time of the Elevatogram is slower than DUET, however it can accurately estimate delays that are 100-times greater than DUET. Source separation algorithms that use time-frequency relative delay estimators may be extended to function with maximum inter-sensor spacings of greater than 2.5cm.

Introduction: Time Delay Estimation (TDE) of a signal is often the first step in multi-channel demixing and source localization, for example TIFROM [1], DUET [2], DEMIX [3] and D-AdRess [4]. With the advent of 5G, the application of Source Separation (SS) in the area of Assisted-Living [5] is increasingly viable. It can be used to enhance a signal of interest to aid social interaction. TDE can be classified into two categories: firstly, Time-Of-Arrival (TOA) [6]; and secondly, Time-Difference-Of-Arrival (TDOA) [7]. Demixing a target signal from a stereo-mixture is a challenge when it experiences a large relative delay. Phase wraparound becomes significant when the distance between the sensors is large. Large inter-sensor separation occurs when sensors are deployed in arbitrary locations, which is a feature of Assisted-Living deployments.

A framework for power weighted relative delay estimators [2] that allows an arbitrary power-weighted estimator to be obtained by selecting the appropriate parameter from a weighted Bregman divergence was provided in [8]. The DUET estimator parametrization [2] yielded the best results in an evaluation of a wide class of estimators in [8]. DUET is used as a benchmark in this paper. Given this evidence in support of DUET's estimators, recent contributions have focused on how to extract overlapping sources in Time-Frequency (TF) bins by using multiple linear spatial filters [9] and on how to select TF bins to increase robustness to noise [10]. The Elevatogram [11] is a recent contribution to the TDE literature. Its advantages have not been fully explored. We contribute an evaluation of the large TDE capability of the Elevatogram and compare it with best-in-class SS TDE [2, 4, 8]. The approach in [4] takes a brute-force approach to relative delay estimation. Its exhaustive search approach make it a good benchmark, yet it is susceptible to phase wraparound and has an unfeasibly high computational load for real-time applications. This study indicates that the Elevatogram is more accurate over a larger range of delays than DUET and D-AdRess. Given a sampling rate of 16 kHz, the Elevatogram estimates relative delays, δ, of speech signals in the range of $|\delta| \leq 800$ samples. It demonstrates the Elevatogram's computational efficiency.

In Figure 1, we consider a stereo anechoic demixing scenario, consisting of J sources, $s_1[n], s_2[n], \ldots, s_j[n], \ldots, s_J[n]$, where $x_1[n] = s_j[n]$ and $x_2[n] = \sum_{j=1}^{J} \alpha_j s_j[n-\delta_j]$, where $j \in \mathbb{Z}_+$. The discrete time index is in the range $1 \leq n \leq N$. We assume that the first mixture $x_1[n]$ contains a good approximation of a single source, $s_j[n]$, due to its physical proximity to one sensor. The second mixture, $x_2[n]$, is the sum of J signals. Each signal is attenuated and delayed by α_j and δ_j. The signal, $s_j[n]$, observed at $x_1[n]$ is present in the mixture $x_2[n]$. The TF transform of any discrete-time signal $s_j[n]$ provides the mapping $\mathbf{S}_j : s_j[n] \in \mathbb{R} \longmapsto \mathbf{S}_j[k,\tau] \in \mathbb{C}$. The discrete frequency is k, where $1 \leq k \leq K$ and τ is discrete time, where $1 \leq \tau \leq T$. A popular choice for multi-channel signal TF analysis is the discrete synchronized-Short-Time Fourier Transform [12]. The TF transform of $x_1[n]$ is denoted $\mathbf{X}_1 \in \mathbb{C}^{K \times T}$. Phase wraparound causes the higher frequency components of the TF representation to be corrupted, making SS difficult.

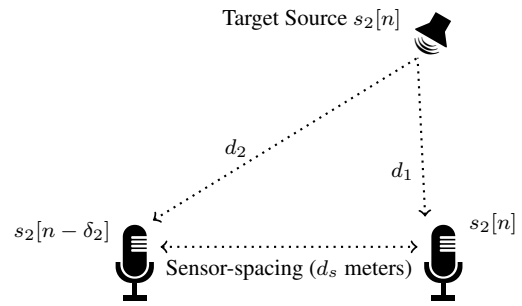

Fig. 1: Physical setup: The paths between the target and mics 1 and 2 are d_1 and d_2, $d_2 > d_1$. The relative delay suffered by the speech signal $s_2[n]$ on d_2 is δ_2 samples relative to the path d_1. As the inter-sensor spacing, d_s, increases, $s_2[n]$ experiences greater phase wraparound in the TF domain.

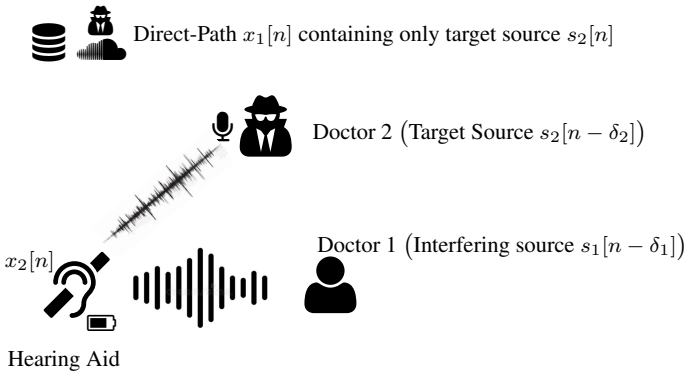

Fig. 2: Scenario: The first mixture, $x_1[n]$, is captured on a cloud-server. It contains the target source $s_2[n]$. The second mixture, $x_2[n]$, is heard by a hearing-aid. It captures a mixture of the two speech signals $s_1[n]$ and $s_2[n]$. We estimate the relative delay, δ_2, experienced bt the target source, s_2[n].

Phase Wraparound: We describe how inter-sensor spacing relates to relative delay and the level of phase wraparound experienced. Figure 1 depicts sensors with a large inter-sensor spacing. This causes large relative delays of sources in the mixtures. Phase wraparound happens because the higher frequency components of the signal are not able to diffract around the microphones located in its path of propagation [13]. These *corrupted* frequencies pose a challenge when using the TF representation of the signal. Larger inter-sensor spacing causes more of the higher frequencies to be corrupted. The bound on the inter-sensor separation in Figure 1, that ensures phase wraparound does not occur is

$$d_s < \frac{\lambda_{min}}{2}, \tag{1}$$

where λ_{min} is the smallest wavelength present in the signal. A relative delay of δ samples for $x_1[n]$ can be expressed in the TF domain as $\mathbf{X}_1[k,\tau] = \mathbf{X}_1[k,\tau]e^{-j\Omega_k\delta}$. The discrete angular frequency is $\Omega_k = \frac{2\pi}{K}k$ for the k^{th} frequency index of a K-point DFT. The phase spectrum of $x_1[n]$ is given by $\angle\mathbf{X}_1[k,\tau] = \tan^{-1}(\mathbf{X}_1[k,\tau])$. Due to phase wraparound, the value in some TF bins is greater than π, $|\Omega_k\delta| > \pi$. It is difficult to compute the true relative delay using phase estimates from multiple frequency bins. We only have the principal phase value, which is bounded by

$$-\pi < \angle\mathbf{X}_1[k,\tau] \leq \pi. \tag{2}$$

The true phase can be larger than π. An abrupt phase wraparound of 2π is observed when the true phase exceeds these bounds [14]. To address the challenge of delay estimation from phase estimates which experience phase wraparound, the Elevatogram [11] was introduced as a way to unwrap the phase to yield accurate estimates of large relative delays.

Motivation: Let us assume a scenario where two doctors are speaking concurrently. In Figure 2 Doctor 2 is wearing a lapel microphone. This microphone enables Doctor 2's voice, $s_2[n]$, to be streamed to the device that performs source localization to assist the assisted living patient. Relative delay estimation may be used here to localize Doctor 2 to facilitate patient interaction with Doctor 2. The first mixture, $x_1[n]$ is

provided via a cloud server which contains Doctor 2's speech, $s_2[n]$. The second mixture $x_2[n]$, observed by a hearing aid for example, is the sum of two signals, the speech of Doctor 1 and 2. These signals experience relative delays of δ_1 and δ_2 samples. The relative attenuation coefficients of the speech signals are α_1 and α_2 in the mixture $x_2[n]$. The problem considered by this letter is how to estimate the relative delay, δ_2, when the inter-sensor spacing is large.

Benchmark Methods: DUET is a binary masking SS technique [2]. Its low computational complexity make it an excellent candidate for relative delay estimation however its range of applicability is limited by the requirement for sensors to be close together, typically $d_s < 2.5$cm. It assumes speech is windowed-disjoint orthogonal [15], which means TF bins can be assigned to one of the sources. This assumption is sufficiently true for appropriate parametrisation of the TF plane. DUET uses attenuation-delay estimates, (α, δ), which are computed using

$$\alpha(k,\tau) = \Big|\frac{\mathbf{X}_2[k,\tau]}{\mathbf{X}_1[k,\tau]}\Big| \text{ and } \delta(k,\tau) = -\frac{1}{\Omega}\angle\frac{\mathbf{X}_2[k,\tau]}{\mathbf{X}_1[k,\tau]}, \tag{3}$$

to construct a power weighted 2-D histogram. Peaks form at the attenuation-delay coordinates of each source in this histogram. D-AdRess [4] was recently introduced to extend instantaneous demixing to an anechoic scenario. D-AdRess uses abrute-force search over a range of delays $1 \leq \delta \leq D_{max}$ as well as attenuations, $0 < \alpha_j \leq 1$, so that relative delays and attenuations are considered in separation. D-AdRess develops peaks at the estimated delay location. A disadvantage of D-AdRess is its large computation time. Figure 2 illustrates the estimates achieved by these benchmark methods for a true relative delay of $\delta_2 = 3$ samples. DUET computes the estimate $\delta_{est} = 2.79$ in Figure 3. D-AdRess yields the estimate $\delta_{est} = 3$ in Figure 4.

Elevatogram Delay Estimation Technique: Let us consider the two source, two mixture scenario shown in Figure 2. The mixtures are $x_1[n]$ and $x_2[n]$. They are denoted $\mathbf{X}_1$ and $\mathbf{X}_2$ in TF. We multiply the first mixture by the complex conjugate of the second mixture element-wise

$$\hat{\mathbf{X}} = \mathbf{X}_1 \odot \overline{\mathbf{X}}_2, \tag{4}$$

where $\overline{\mathbf{X}}_2$ is the complex conjugate of $\mathbf{X}_2$. We quantize the phase, $\phi = |\Omega_k \delta| < \pi$, into L uniformly spaced levels. For the k-th row in $\angle\hat{\mathbf{X}}[k,:]$, we construct a histogram, $\mathbf{h}_k$, to categorize the phase content by counting the same phase measurements in relevant bins using the operator

$$\mathbf{h}_k = \text{hist}\left(\angle\hat{\mathbf{X}}[k,:], L\right). \tag{5}$$

The L-levels correspond to the bins of the histogram, $\mathbf{h}_k^T$. The taller the count within a bin, the more intensely that particular bin has been activated. We do this for all the available frequencies, K. The result is a phase-frequency matrix, called the Elevatogram

$$\mathbf{P} = [\mathbf{h}_1^T, \mathbf{h}_2^T, \ldots \mathbf{h}_K^T] \in \mathbb{R}^{L \times K}, \tag{6}$$

which is depicted in Figure 5. In this worked example, it consists of slanting lines, exhibiting phase wraparound of 2π at 2000 and 7000Hz. The larger the delay, the more often this phase wraparound is observed. The three most visible lines in Figure 5 are a single line that undergoes phase wraparound. We determine the location of the set of the most significant collinear points forming a straight line, as shown in Figure 5. In the traditional voting procedure, inspired by the classical Hough transform [16], we parameterize the matrix, $\mathbf{P}$, by distance-angle, (ρ, ϕ), to determine the value of the accumulator cell that receives the highest vote in the accumulator. This is displayed as the brightest point in Figure 6 which is the global maximum (ρ_{max}, ϕ_{max}) of the accumulator. We obtain ϕ_{max} and calculate the estimated delay

$$\delta_{est} = -\frac{K}{L}\tan\left(\phi_{max}\right). \tag{7}$$

Evaluation Set-up: Experiments are conducted using real speech utterances from the TIMIT corpus [17], which have a sampling-rate of $F_s = 16$ kHz. A target, ground-truth delay is used to generate anechoic mixtures. A K-sample FFT Hamming window is used, where $K = 1024$ or 2048 samples and the number of quantization phase-levels is $L = 100$ or 50, depending on the size of the delay to be estimated. We use Equation (7) as a guide to selecting parameters, using the relations $\delta_{est} \propto K$ and $\delta_{est} \propto \frac{1}{L}$. Empirically adjusting K and L gives us a good estimate of δ_{est}. We evaluate each algorithm over the delay range of 3

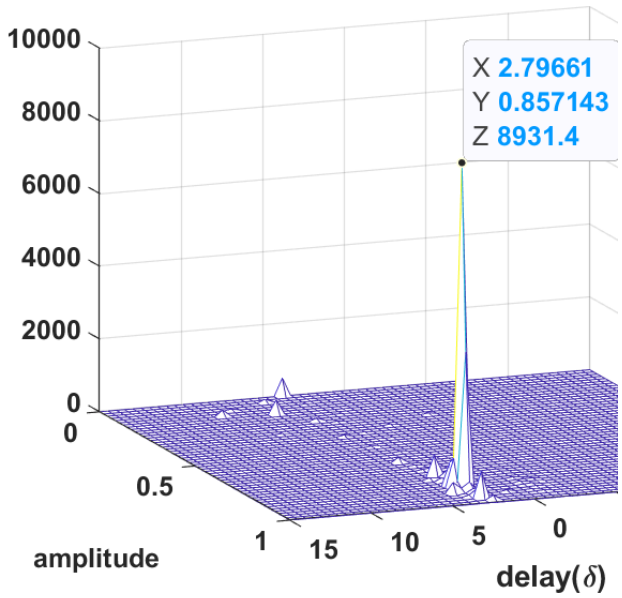

Fig. 3: The ground-truth relative delay is $\delta_2 = 3$ samples. In DUET, a peak forms which yields an estimate of $\delta_{est} = 2.79$ samples. The relative attenuation estimate is $\alpha_2 \approx 0.8$.

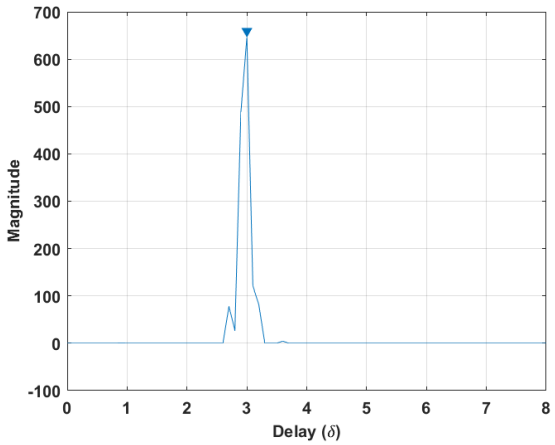

Fig. 4: D-AdRess produces an estimate of $\delta_{est} = 3$ samples.

up to 800 samples, and analyze how close, δ_{est}, is to the ground-truth, δ_2, using Matlab R2021a and a 16GB, Intel(R) Core(TM) i7-6500U CPU@2.5GHz, 2.6 GHz, 64-bit operating system, x64-based processor using 1000 Monte Carlo trials per mixing scenario.

Results: Firstly, we evaluate the Elevatogram in three settings to demonstrate its range of applicability and to show how the target delay impacts the section of L and K. The delay estimation problem is categorized into small-delay, medium-delay and large-delay estimation. Secondly, we examine the largest ground-truth delay, δ_2, that can be estimated. We benchmark the Elevatogram against DUET and D-AdRess in terms of accuracy, speed and robustness using 1000 Monte Carlo trials.
Small Delay: Let us assume that the ground-truth value is $\delta_2 = 3$ samples. We use a $K = 1024$-point FFT and the number of quantization levels is $L = 100$ in the Elevatogram. In Figure 6, the brightest point ϕ_{max} appears at $\phi = 2.86$ rads. Substituting these values in Equation (7) gives us $\delta_{est} = \frac{1024}{100} \times \tan\left(2.86\right) \approx 3$ samples, which closely approximates the true value.
Medium delay: For a ground-truth delay of $\delta_2 = 100$ samples, we increase the FFT size to $K = 2048$ and the quantization level is kept the same at $L = 100$. The counterpart of the accumulator in Figure 6 gives an estimate of $\phi = 1.7709$ rads. Substituting these values into Equation (7) gives us $\delta_{est} = -\frac{2048}{100} \times \tan\left(1.7709\right) \approx 100.97$ samples, which implies that the error is 0.97 samples.

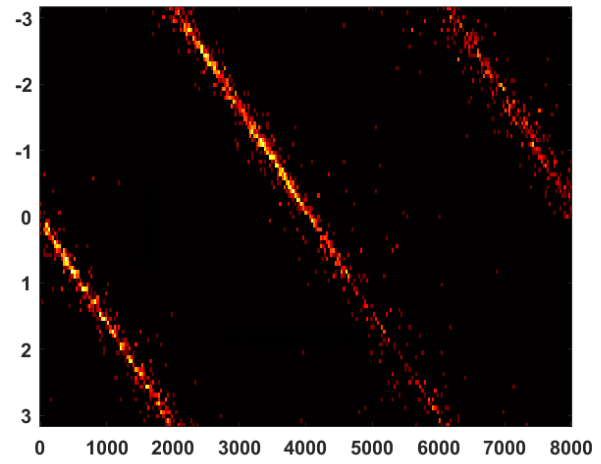

Fig. 5: Elevatogram: Slanting lines are observed for real speech utterances. They indicate the energy concentration of a particular phase as a function of frequency. Phase wraparound is observed when lines hit the boundary at frequencies 2000 Hz and 6000 Hz. The larger the delay, the more often phase wraparound occurs.

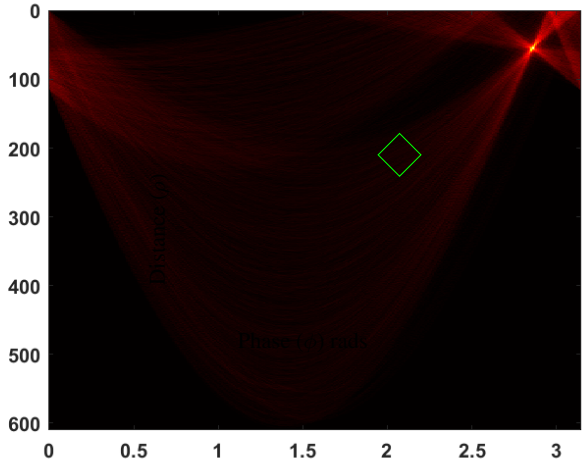

Fig. 6: Distance(ρ)-Phase(ϕ) accumulator: The voting algorithm gives us an estimate of the brightest point (at $\phi = 2.5$ rads), which is used to estimate the delay δ_2.

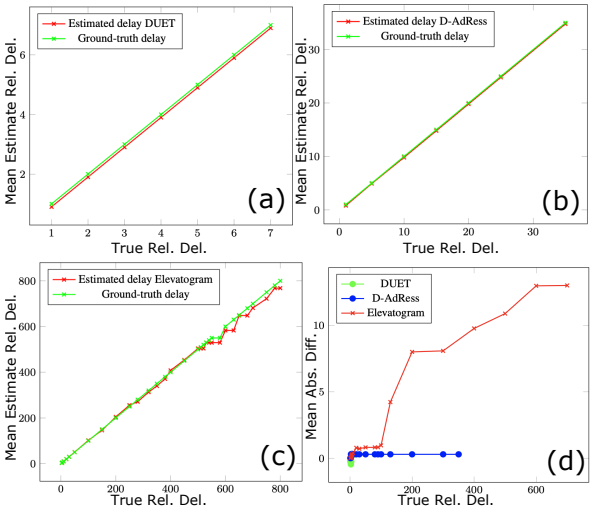

Fig. 7: Delay estimates (in samples) for speech from the TIMIT database: **(a)** DUET can estimate delays of up to 7 samples. Peaks are difficult to detect for big delays due to phase wraparound. **(b)** D-AdRess accurately estimates delays of up to 30 samples under similar conditions. Its execution time depends on the delay to be estimated. **(c)** The Elevatogram can estimate delays of up to 800 samples. As delays become larger, greater than 500 samples, the estimated delays (red) estimate the true delay value (green) less accurately. **(d)** The Elevatogram is significantly better in measuring large delays and produces estimates which have a small Mean Absolute Difference.

Large delay: In the final case the ground-truth delay is $\delta_2 = 500$ samples. We keep the $K = 2048$-point Hamming window but decrease the number of quantization levels to $L = 50$. The counterpart of the accumulator in Figure 6, gives a ϕ_{max} at approximately $\phi = 1.6508$ rads. Substituting these values into Equation (7) gives us $\delta_{est} = -\frac{2048}{50} \times \tan\left(1.6508\right) \approx 510.88$ samples. The error in this case is 10.88 samples.

We bench-mark the Elevatogram against DUET and D-AdRess for speech signals. We investigate if the Elevatogram is significantly better than DUET and D-AdRess for large delay estimation. In Figure 7, DUET estimates delays up to $\delta_2 = 7$ samples accurately. Peak formation at the appropriate delay location on the attenuation-delay plane is not always guaranteed. On the other hand, D-AdRess is precise in estimating delays.

Figure 7 depicts how D-AdRess performs at measuring medium delays, up to 35 samples, under similar conditions. Figure 8 depicts that the delay range parameter of D-AdRess must increase as the delay value to be estimated increases. A consequence of this is an increase in the Matlab execution time of D-AdRess. In Figure 7, the Elevatogram approach surpasses the large delay estimation ability of both DUET and D-AdRess by a large margin. The Elevatogram successfully estimates large delays of up to 800 samples. After 550 samples, the estimated delay, δ_{est} tends to give poorer estimates of the true delay value. In panel (d) of Figure 7 we illustrate the robustness of the estimators by plotting the Mean Absolute Difference (MAD) between the estimated relative delay and true relative delay, as a function of delay. Due to phase wraparound and the intersensor spacing, the range of relative delays that can be estimated using DUET without phase wraparound is small. Consequently, DUET and D-AdRess give small MAD for small, true relative delays. In comparison, the MAD of the Elevatogram is small, even for relative delays of up to 550 samples. A 10-sample MAD for a true relative delay of 500 samples, implies that on average the error in the estimate is 2%. In summary, DUET, though more computationally efficient than Elevatogram is inappropriate for large delay estimation of speech signals due to phase wraparound. Figure 9 depicts that D-AdRess is computationally inefficient, as it takes around 70s to estimate a delay of $\delta_2 = 2$ samples. The Matlab execution time of both DUET and the Elevatogram does not depend on the delay. Given that the Elevatogram's computation time does not depend on delay, its robustness and its ability to accurately estimate large delays, it is a preferable estimator to DUET and D-AdRess.

Conclusion: We compared a large delay estimation technique known as the Elevatogram with DUET and D-AdRess. We observed that the Elevatogram can estimate large delays of up to 800 samples for speech. The largest delays DUET and D-AdRess estimated were 7 samples and 35 samples, respectively. D-AdRess is computationally inefficient. Its computational time in directly proportional to the delay that is to be estimated. This involves a brute-force technique to iterate over the two ranges, an attenuation range of size A, and a delay range of size D. This process is repeated over all STFT Hamming windows, T. In summary, it imposes an additional computational complexity of $O(A \times D \times T)$ FLOPS. A K-point FFT costs $O(K \log(K))$ FLOPS. This accounts for the greater execution time of D-AdRess. To measure a delay of 7 samples, D-AdRess runs in 120s in Matlab, whereas the Elevatogram and DUET run in 25s and $\approx$ 1s, respectively. We express our results in the context of a real-world deployment. We considered real TIMIT speech utterances of 16 kHz sampling rate. Estimating a large delay of 550 samples corresponds to a delay in seconds of $t = 0.034$s. Given that sound travels in dry air at a speed of $s = 331.29$ m/s, if we use the Elevatogram as part of a SS algorithm, we will be able to perform speech localization and separation in controlled Robotic Auditory room [18] of where delays can correspond to path distances of up to $d = 11.38$m.

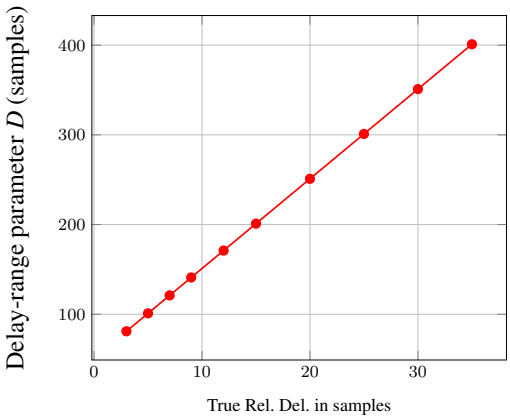

Fig. 8: D-AdRess uses a delay-range parameter that must increase as the delay value to be estimated increases.

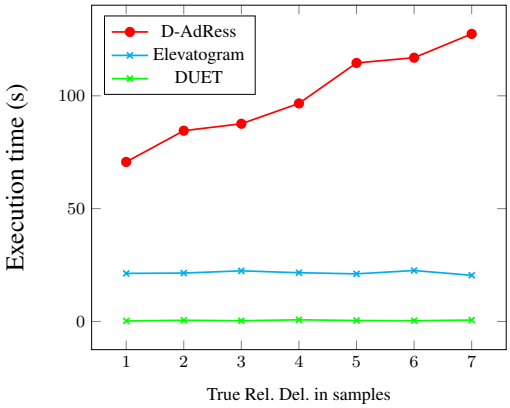

Fig. 9: D-AdRess' execution time becomes large as the delay-range parameter increases. A comparison of DUET, D-AdRess and Elevatogram with regard to execution time indicates that the fastest algorithm is DUET (it runs in 0.2s) and the slowest is D-AdRess. For delays of $\delta_2 = 7$ samples, D-AdRess takes approximately 120 seconds. The Elevatogram takes 20 seconds to execute.

Acknowledgment: This paper has emanated from research supported in part by a Grant from Science Foundation Ireland under Grant number 18/CRT/6222, 13/RC/2077_P2 and 15/SIRG/3459

Swarnadeep Bagchi and Ruairí de Fréin (*TU Dublin, Ireland*)
E-mail: D18128352@mytudublin.ie and ruairi.defrein@tudublin.ie

References

1 F. Abrard and Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources," *Sig. Proc.*, vol. 85, no. 7, pp. 1389–1403, 2005.
2 O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Sig. Proc.*, vol. 52, no. 7, pp. 1830–1847, 2004.
3 S. Arberet, R. Gribonval, and F. Bimbot, "A robust method to count and locate audio sources in a multichannel underdetermined mixture," *IEEE Trans. Sig. Proc.*, vol. 58, no. 1, pp. 121–133, 2009.
4 S. Bagchi and R. de Fréin, "Extending instantaneous de-mixing algorithms to anechoic mixtures," in *IEEE ISSC*, 2021, pp. 1–6.
5 K. Witrisal, P. Meissner, E. Leitinger, Y. Shen, C. Gustafson, F. Tufvesson, K. Haneda, D. Dardari, A. F. Molisch, A. Conti *et al.*, "High-accuracy localization for assisted living: 5G systems will turn multipath channels from foe to friend," *Sig. Proc. Mag.*, vol. 33, no. 2, pp. 59–70, 2016.
6 R. Wu, J. Li, and Z.-S. Liu, "Super resolution time delay estimation via mode-wrelax," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 1, pp. 294–307, 1999.
7 G. Carter, "Time delay estimation for passive sonar signal processing," *IEEE Trans. ASSP*, vol. 29, no. 3, pp. 463–470, 1981.
8 R. de Fréin and S. T. Rickard, "Power-weighted divergences for relative attenuation and delay estimation," *IEEE Sig. Proc. Ltrs*, vol. 23, no. 11, pp. 1612–1616, 2016.
9 Y. He, H. Wang, Q. Chen, and R. H. So, "Harvesting partially-disjoint time-frequency information for improving degenerate unmixing estimation technique," in *ICASSP*, 2022, pp. 506–510.
10 J. Geng, S. Wang, S. Gao, Q. Liu, and X. Lou, "A time-frequency bins selection pipeline for direction-of-arrival estimation using a single acoustic vector sensor," *IEEE Sen. J.*, vol. 22, no. 14, pp. 14 306–19, 2022.
11 R. de Fréin, "Tiled time delay estimation in mobile cloud computing environments," in *IEEE ISSPIT*, 2017, pp. 282–287.
12 R. de Fréin and S. T. Rickard, "The synchronized short-time-Fourier-transform: properties and definitions for multichannel source separation," *IEEE Trans. Sig. Proc.*, vol. 59, no. 1, pp. 91–103, 2010.
13 B. C. Moore, *An introduction to the psychology of hearing*. Brill, 2012.
14 P. Mowlaee, J. Kulmer, J. Stahl, and F. Mayer, *Single channel phase-aware signal processing in speech communication: theory and practice*. John Wiley & Sons, 2016.
15 S. Rickard and O. Yilmaz, "On the approximate w-disjoint orthogonality of speech," in *IEEE ICASSP*, vol. 1, 2002, pp. I–529.
16 R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Comm. ACM*, vol. 15, no. 1, pp. 11–15, 1972.
17 J. S. Garofolo, "TIMIT acoustic phonetic continuous speech corpus," *Linguistic Data Consortium, 1993*, 1993.
18 H. Sawada and C. Udaka, "A robotic auditory system for imitating human listening behavior," in *ICMA*, 2013, pp. 773–778.