



П. І. Рожак, Н. О. Семенишин

ЗАСТОСУВАННЯ SRP-МЕТОДУ ДЛЯ ЛОКАЛІЗАЦІЇ ДЖЕРЕЛА АКУСТИЧНОГО СИГНАЛУ В ОДНОВИМІРНІЙ ОБЛАСТІ

Локалізація джерела звуку за допомогою цифрового оброблення сигналу потрібна для того, щоб оцінити напрямок, звідки приходить конкретний акустичний сигнал, а це, водночас, важливо для реалізації "hand-free" систем та систем позиціонування, наприклад у відео конференціях. У будь-якому разі потрібен алгоритм, який дає змогу надійно локалізувати розташування акустичного джерела звуку. У ролі такого алгоритму у цій роботі реалізовано SRP-метод із фазовою трансформацією. Перевагами SRP-методу є точність знаходження джерела у середовищах з реверберацією та шумом, а також малий часовий відрізок вхідного сигналу для коректної роботи. Цей алгоритм аналізує потужність звуку, отриманого мікрофонною решіткою, у певній точці просторової сітки деякого приміщення. Тобто замість оцінки набору тимчасових різниць часу прибуття між парами мікрофонів і об'єднанням отриманих оцінок, щоб знайти місце розташування джерела, за непрямыми методами пошук розташування джерела працює перебором просторових точок на сітці досліджуваної області. Представлено кілька експериментів із двома мікрофонами для одновимірної задачі локалізації звуку у довгому приміщенні, ширина якого нехтувалася. Ефективність знаходження акустичного джерела досліджено на мовному сигналі та довільному – плеск в долоні. Показано точність цього методу до кількох сантиметрів навіть за наявності деякого шуму та реверберації у вхідному сигналі.

Ключові слова: SRP-метод (Steered Response Power); фазова трансформація; швидке перетворення Фур'є (ШПФ); взаємна кореляція; частота дискретизації; аудіоінтерфейс.

Вступ. Методи локалізації джерела звуку та поєднання мікрофонних решіток і оброблення сигналів на сьогодні активно застосовують для отримання високоякісного мовного сигналу. Такі застосування використовують просторову фільтрацію, що дає змогу підвищити рівень сигналу одного мовця, і погасити інші сигнали.

Тоді, як деякі системи призначені для того, щоб зосередитися на звуках, які виходять із заданого місця розташування, більшість використовує адаптивні алгоритми, які відстежують розташування одного або кількох мовців і регулюють фокус масиву сенсорів відповідно. Це "електронне керування" усуває необхідність обладнання ручного керування. Окрім цього, ці системи мають великі перспективи у розробленні так званих розумних кімнат (*smart rooms*), які активно взаємодіють із людиною за допомогою багатьох сенсорів.

Зважаючи на широке розмаїття сфер застосування системи локалізації джерела звуку, така програмно-технічна система є дуже актуальною та корисною на сьогодні.

Опис SRP-методу. У цій роботі використано SRP-метод, який належить до класу непрямих методів локалізації акустичних джерел. Замість оцінки набору тимчасових різниць часу прибуття (time difference of arrival – TDOA) між парами мікрофонів і об'єднанням отриманих оцінок, щоб знайти місце розташування джерела, за непрямыми методами, пошук розташування джерела працює перебором просторових точок на сітці досліджуваної області. Метод є стійким до шуму і реверберації (Belloch et al., 2013).

Для часового інтервалу n , SRP, $P_n(\vec{x})$, є функціонал дійсних значень для просторового вектора \vec{x} , отриманого зміною фазових затримок вхідних сигналів та сумуванням їх для усіх мікрофонів. За гіпотезою максимум $P_n(\vec{x})$ вказує на місцезнаходження k джерел звуку $\hat{x}_s^{(n)}(k)$ навіть за значного шумового фону та реверберації. Найбільші максимуми $P_n(\vec{x})$ формують набір $\hat{x}_s^{(n)}(k)$. Наприклад, для одиничного джерела звуку, оцінка розташування становить

$$\hat{x}_s^{(n)}(1) = \arg \max_{\vec{x}} P_n(\vec{x}). \quad (1)$$

Нехай дано сигнал $m_i(t)$, що є виходом мікрофона i у мікрофонній решітці розміром M . Тоді оцінка SRP для інтервалу скінченної довжини T визначена

$$P_n(\vec{x}) \equiv \int_{nT}^{(n+1)T} \left| \sum_{i=1}^M w_i m_i(t - \tau(\vec{x}, i)) \right|^2 dt, \quad (2)$$

де w_i – деякий ваговий коефіцієнт і $\tau(\vec{x}, i)$ – час проходження сигналу з точки \vec{x} до мікрофона i . Значення SRP може бути точно розраховано сумуванням узагальненої взаємної кореляції для всіх можливих пар мікрофонів (DiBiase, 2000), а як відомо, на основі теореми про згортку дискретного перетворення Фур'є (ДПФ), кореляцію двох сигналів у часовій області, можна представити як добуток цих сигналів у частотній області отриманих за допомогою ДПФ (Babak, Khandetskiy, &

Інформація про авторів:

Рожак Петро Ігорович, аспірант, асистент кафедри інформаційних технологій. **Email:** petrrozhak@gmail.com

Семенишин Назар Олегович, аспірант, інженер кафедри інформаційних технологій. **Email:** xa4abu@ukr.net

Цитування за ДСТУ: Рожак П. І., Семенишин Н. О. Застосування srp-методу для локалізації джерела акустичного сигналу в одновимірній області. Науковий вісник НЛТУ України. 2017. Вип. 27(5). С. 160–163.

Citation APA: Rozhak, P. I., & Semenyshyn, N. O. (2017). SRP-method for locating an acoustic signal source in a one-dimensional region. Scientific Bulletin of UNFU, 27(5), 160–163. <https://doi.org/10.15421/40270531>

Shriufer, 1996; Nussbaumer, 1985). При переході (2) у частотну область, використовуючи зважені коефіцієнти $W_l^*(\omega)$ та застосовуючи теорему Парсеваля, отримуємо

$$P_n(\bar{x}) = \sum_{k=1}^M \sum_{l=k+1}^M \int_{-\infty}^{\infty} W_k(\omega) W_l^*(\omega) M_k(\omega) M_l^*(\omega) e^{j\omega \tau_{kl}(\bar{x})} d\omega. \quad (3)$$

Визначимо вагову функцію як

$$\Psi_{kl}(\omega) \equiv W_k(\omega) W_l^*(\omega). \quad (4)$$

Перепишемо вираз (3) з урахуванням симетрії елементів суми

$$P_n(\bar{x}) = \sum_{k=1}^M \sum_{l=k+1}^M \int_{-\infty}^{\infty} \Psi_{kl}(\omega) M_k(\omega) M_l^*(\omega) e^{j\omega \tau_{kl}(\bar{x})} d\omega. \quad (5)$$

Фазове перетворення (phase transform – PHAT), особливо ефективне при визначенні взаємної кореляції для знаходження різниці часу прибуття з мовного сигналу при шумовому фоні та реверберації (Hoang Do et al., 2007; DiBiase, 2000). Визначимо вагові значення як обернену величину частотних складових

$$\Psi_{kl}(\omega) \equiv \frac{1}{|M_k(\omega) M_l^*(\omega)|}. \quad (6)$$

Наведемо визначення параметра

$$\tau_{kl}(x) = \frac{\|x - x_k\| - \|x - x_l\|}{c}. \quad (7)$$

Ця функція представляє теоретичне значення затримки надходження звуку від джерела, розташованого у точці x , для пари мікрофонів (k, l). Коефіцієнт c – значення швидкості звуку у повітрі.

Отже, розв'язком нашої задачі є знаходження одного або кількох значень максимумів $\hat{x}_s^{(n)}(k)$ шляхом дослідження поверхні $P_n(\bar{x})$ по всій досліджуваній області. Розрахунок будь-якої конкретної точки x для $P_n(\bar{x})$ називають *оцінкою функціоналу*. Для SRP-PHAT функціоналу (5), хочемо визначити точку-джерело місце в кімнаті, що дає максимальне значення. Для оптимізації функціоналу можемо застосувати метод Контракції сто-

хастичної області (stochastic region contraction – SRC) (Hoang Do et al., 2007).

Результати проведених експериментів. У цій роботі здійснено кілька експериментів із двома мікрофонами у вузькому коридорному приміщенні довжиною 6 м, де ширина приміщення нехтувалась і розрахунки велись тільки по одній координаті довжини.

Звуки, які записували, були різної природи: слово "Раз" та плеск в долоні. Виміри здійснювали на різних відстанях від мікрофонів. У коридорному приміщенні при записі вхідного сигналу виникало ще невелике ехо та шум з вулиці крізь відчинені вікна. Для всіх дослідів частота дискретизації дорівнює 48000 Герц, а один відлік звукового сигналу для одного каналу займає 24 біти.

Для записування звуку використовували звичайні динамічні мікрофони, розташовані за наперед визначеними координатами на кінцях приміщення. Величина по координаті x рахується зліва направо, від лівого каналу до правого, і при $x=6$ джерело звуку знаходиться біля мікрофона, який відповідає за правий канал.

Для синхронного запису кількох звукових доріжок (з мінімальними затримками) використовуємо аудіо-інтерфейс Roland ediol ua-25 із двома каналами входу. Щоб здійснити вибірку фрагмента сигналу для кожної доріжки використати вільне ПЗ Audacity. Розмір фрагмента вибірки для формування вхідного сигналу для всіх експериментів становить 8192 (прибл. 170 мс).

Експеримент 1. Сигналом є плеск в долоні. На рис. 1 представлено вже відібрані фрагменти вхідного сигналу. Розташування джерела звуку по осі x становить 5 м 60 см. Поверхню функціоналу $P_n(\bar{x})$ із формули (5) представлено на рис. 2. Максимум цієї поверхні розташований у точці $x=5,54$ м, що становить похибку шість сантиметрів щодо справжнього розташування джерела звуку.

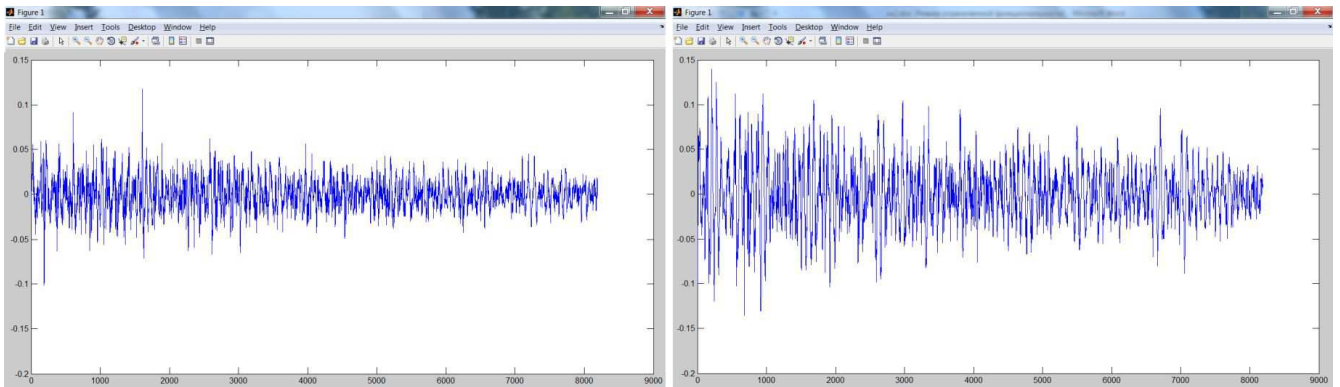


Рис. 1. Плеск в долоні. Вигляд звукового сигналу для лівого (зліва) і правого (справа) каналів

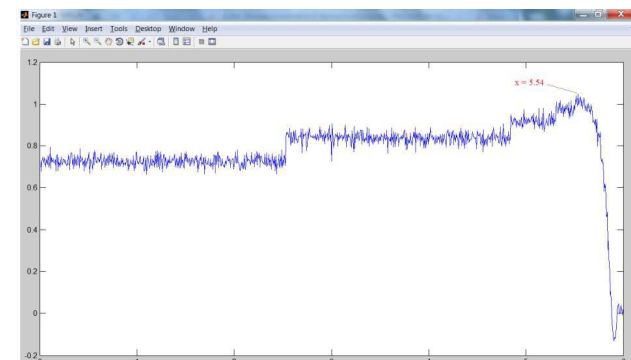


Рис. 2. Вигляд $P_n(\bar{x})$ для експерименту 1 в одновимірній області

Експеримент 2. Вхідним звуком цього разу є слово "Раз". Запис здійснюється безпосередньо у мікрофон лівого каналу з максимально близької відстані. На рис. 3 різним кольором показано дві звукові доріжки цього сигналу, з яких буде здійснюватись вибірка фрагментів для алгоритму локалізації. Поверхню функціоналу $P_n(\bar{x})$ із формули (5) представлено на рис. 4.

Тут чітко видно, що диктор знаходиться на лівому кінці приміщення на дуже близькій відстані до мікрофону при $x=0,04$.

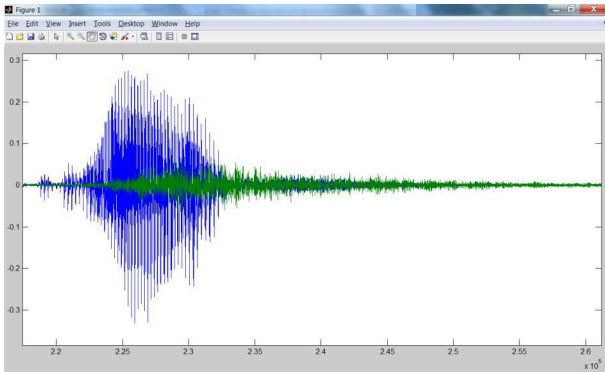


Рис. 3. Слово "Раз". Темніший колір більшої амплітуди – запис з лівого каналу

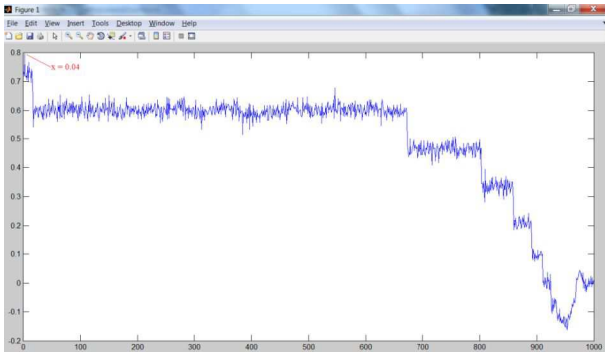


Рис. 4. Вигляд $P_n(\bar{x})$ для експерименту 2 в одновимірній області

Висновки. У цьому дослідженні застосовано SRP-метод із фазовою трансформацією для задачі локалізації джерела звуку. Замість оцінки набору тимчасових різниць часу прибуття між парами мікрофонів і об'єднанням отриманих оцінок, для знаходження місця

розташування джерела непрямі методи, тобто перебір просторових точок на сітці досліджуваної області.

Алгоритм є менш чутливим до реверберацій та шумів вхідного звуку за рахунок множення на вагові коефіцієнти, тобто при так званій фазовій трансформації. Також до переваг алгоритму можна віднести можливість додавання великої кількості мікрофонів, локалізацію кількох рухомих об'єктів та адаптованість для зміни оптимізаційного методу, щоб досягнути більшої швидкодії.

Здійснено низку експериментів із двома мікрофонами у вузькому коридорному приміщенні, де ширина приміщення нехтувалась і розрахунки велись тільки по одній координаті довжини. Середня похибка відхилення реального розташування джерела звуку від того, яке оцінюється функціоналом, становила 1-2 %.

Перелік використаних джерел

- Babak, V. P., Khandetskiy, V. S., & Shriufier, E. P. (1996). *Obrobka syhnaliv*. Kyiv: Lybid, 390 p. [in Ukrainian].
- Belloch, J. A., Gonzalez, A., Vidal, A. M., Cobos, M. (2013). Real-time Sound Source Localization on Graphics Processing Units. *Procedia Computer Science*, 18, 2549–2552. <https://doi.org/10.1016/j.procs.2013.05.438>
- DiBiase, J. H. (2000). *A High Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments using Microphone Arrays* (PDF) (PhD.). Providence, Rhode Island, Brown Univ.
- Hoang Do, Harvey F. Silverman, Ying Yu (2007). A Real-Time SRP-PHAT Source Location Implementation using Stochastic Region Contraction(SRC) on a Large-Aperture Microphone Array. *ICASSP 2007. IEEE International Conference: Acoustics, Speech and Signal Processing*. Honolulu, HI, USA. <https://doi.org/10.1109/ICASSP.2007.366631>
- Nussbaumer, G. (1985). *Bystroe preobrazovanie Fure i algoritmy vychisleniia svertok*. Moscow: Radio i sviaz, 248 p. [in Russian].

П. И. Розжак, Н. О. Семенишин

Національний лесотехнічний університет України, г. Львов, Україна

ПРИМЕНЕНИЕ SRP-МЕТОДА ДЛЯ ЛОКАЛИЗАЦИИ ИСТОЧНИКА АКУСТИЧЕСКОГО СИГНАЛА В ОДНОМЕРНОЙ ОБЛАСТИ

Представлен метод определения местонахождения источников звука с использованием сегмента данных из массива микрофонов. Использован надежный метод управления мощностью отклика с использованием фазового преобразования (steered response power using the phase transform – SRP-PHAT). В этом алгоритме истинное местоположение источника имеет максимальное значение функционала SRP-PHAT. Метод очень привлекательный в приложениях акустической локализации из-за его работоспособности в средах с реверберацией. В представленном методе, базируясь на известном алгоритме SRP-PHAT, предложен функционал, основанный на накоплении значений взаимных корреляций для каждой пары микрофонов в диапазоне, который охватывает каждую точку определенной пространственной сетки. Результаты использования алгоритма представлены на двух графиках в одномерной области. В качестве аппаратного обеспечения использованы два микрофона и аудиоинтерфейс, а в качестве программного Audacity – аудиоредактор звуковых файлов, ориентированный на работу с несколькими дорожками. Показано, что результаты локализации акустического источника с использованием этого метода дают хорошую оценку положения источника со средним отклонением относительно точного местоположения источника, равного 1-2 %. Установлено, что записываемый звук может иметь различную природу, например человеческий голос или обычное хлопанье.

Ключевые слова: SRP-метод (Steered Response Power); фазовая трансформация; быстрое преобразование Фурье (БПФ); взаимная корреляция; частота дискретизации; аудиоинтерфейс.

P. I. Rozhak, N. O. Semenyshyn

Ukrainian National Forestry University, Lviv, Ukraine

SRP-METHOD FOR LOCATING AN ACOUSTIC SIGNAL SOURCE IN A ONE-DIMENSIONAL REGION

Signal processing for acoustic sound source localization is required in order to find the direction from where a particular sound source signal is coming and it is also important to find a solution for hands free communication. Hand free communications and video conferencing are different applications requiring sound source localization. In this paper we implement such method for locating sound sources. The method employs the proven-robust steered response power using the phase transform (SRP-PHAT) as a functional one. In this algorithm the true source location has the maximum SRP-PHAT value, which can be found by applying some optimization method with respect to one parameter x. The steered response power phase transform (SRP-PHAT) is a beamformer method very attractive in acoustic localization applications due to its robustness in reverberant environments. Therefore sound source locali-

zation and speech/non-speech detection techniques have been presented in this work to be used in a multiparticipant environment with a microphone system. Based on the well-known SRP-PHAT method we propose functional which is based on the accumulation of generalized cross-correlations values for each microphone pair in a range that covers the volume surrounding each point of the defined spatial grid. During calculation we made fast Fourier transform and use convolution theorem of discrete Fourier transform to get convolution of two signals in time domain by multiplying these signals in frequency domain and then again returning to time domain by using inverse fast Fourier transform. The results of using algorithm are presented on two plots in one dimensional domain. Two microphones and an audio interface were used as hardware and the Audacity as multi-track audio editor and recorder. It is shown, that the results of acoustic source localization using this method give good evaluation of source position with average deviation with respect to exact source location equal 1-2 %. Also it was shown that recording sound can have different nature like human voice or clapping and environment for experiment may contain some level of noise and reverberation.

Keywords: SRP-method (Steered Response Power method); phase transform; fast Fourier transform; mutual correlation; sampling rate; audio interface.