# ЛАБЕЛИРАЊЕ ГОВОРНИХ СЕГМЕНАТА У РЕВЕРБЕРАНТНОЈ ПРОСТОРИЈИ УЗ ПОМОЋ МИКРОФОНСКОГ НИЗА

Зоран Шарић[1], Слободан Јовичић[2]

[1]*Институт Безбедности, Краљице Ане бб, 11000 Београд, sare@yubc.net*
[2]*ЕТФ, Краља Александра 73, 11000 Београд, jovicic@etf.bg.ac.yu*

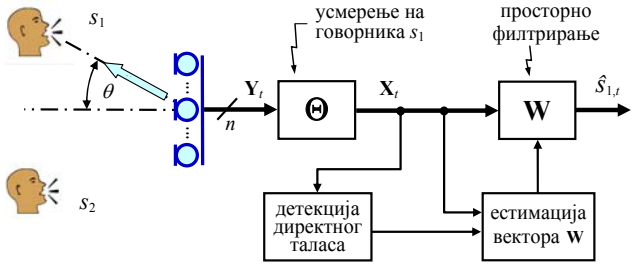
**Садржај** – *У обради сигнала микрофонског низа јавља се потреба за раздвајањем говорних сегмената који потичу од одабраног говорника од сегмената који садрже коктел парти сметње. У раду се предлаже поступак детекције директног таласа одабраног говорника базиран на временском кашњењу рефлексија у односу на директан талас. Критеријум детекције је средњеквадратна грешка отступања од усвојеног модела. Обрада сигнала се врши у временском домену што обезбеђује мало време кашњења детекције. Предложени поступак детекције је тестиран на симулацији просторије са реверберацијом и примењен је на адаптивно филтрирање методом максималне веродостојности. Добијени резултати показују да се применом предложеног поступка детекције остварује бољи квалитет реконструисаног говорног сигнала у односу на уобичајене адаптивне поступке обраде микрофонских сигнала.*

## 1. УВОД

Потреба за раздвајањем говорних сегмента који потичу од одабраног говорника и сегмената који садрже коктел парти сметње проистекао је из примене адаптивног алгоритма минималне варијансе у условима реверберације просторије. Познато је да се у примени алгоритама минималне варијансе у условима реверберације просторије јавља нежељено поништавање корисног сигнала [1],[2]. Једини начин да се избегне ово нежељено поништавање јесте да се издвоје делови сигнала на којима је присутан само користан сигнал и делови сигнала на којима су присутне само сметње и да се на њима примене различите процедуре адаптације [3], [4].

Проблем издвајања делова сигнала са доминантним корисним сигналом и доминантним сметњама обрађено је у више радова [2], [4], и у њима је примењен поступак на бази процене односа сигнал сметња. У овом раду се предлаже другачији поступак посредством детекције директног таласа корисног сигнала. У детекцији директног таласа користи се релативно кашњење рефлексија у односу на директан талас. Критеријум детекције је средњеквадратна грешка отступања од усвојеног модела. Када је средњеквадратна грешка испод прага одлуке, доноси се одлука да је присутан само директан талас. У супротно, сматра се да или нема директног таласа или су поред директног таласа присутне рефлексије таласа и сметње.

Обрада сигнала се врши у временском домену што обезбеђује кратко време кашњења детекције. Предложени поступак детекције је тестиран на симулацији просторије са реверберацијом и примењен на адаптивно филтрирање методом максималне веродостојности. Добијени резултати показују да се применом предложеног поступка детекције остварује бољи квалитет реконструисаног говорног сигнала у односу на уобичајене адаптивне поступке обраде микрофонских сигнала.




Сл. 1. *Блок дијаграм обраде сигнала у просторији са реверберацијом.*

## 2. ПОСТАВКА ПРОБЛЕМА

Модел произвођења микрофонских сигнала дат је релацијом

$$\mathbf{H}\mathbf{S}_t + \mathbf{N}_t = \mathbf{Y}_t \qquad (1)$$

где је $\mathbf{Y}_t$ вектор $n$ микрофонских сигнала, $\mathbf{S}_t$ вектор $m$ извора звука од којих је само прва компонента извор $s_0$ користан сигнал, $\mathbf{H}$ је матрица преноса од сваког узвора звука до сваког микрофона, а $\mathbf{N}_t$ је вектор адитивног шума на микрофонима. Компоненте шума микрофона су међусобно некорелисане те је коваријациона матрица овог вектора једнака $\mathbf{R}_N = E\left(\mathbf{N}\mathbf{N}^*\right) = \sigma^2\mathbf{I}$ .

Блок дијаграм обраде сигнала је приказан на слици 1. Пријемни сноп усмерава се на одабраног говорника компензовањем временских кашњење таласа до сваког од микрофона према следећој релацији.

$$\mathbf{\Theta}\mathbf{Y}_t = \mathbf{\Theta}\mathbf{H}\mathbf{S}_t + \mathbf{\Theta}\mathbf{N}_t = \tilde{\mathbf{H}}\mathbf{S}_t + \tilde{\mathbf{N}}_t = \mathbf{X}_t \quad , \qquad (2)$$

где је $\mathbf{\Theta} = diag(1 \quad q^{-\tau_1} \quad \cdots \quad q^{-\tau_{n-1}})$ матрица компензације кашњења, где је са $q^{-\tau_k}$ означен оператор кашњења за $\tau_k$ . Матрица $\tilde{\mathbf{H}}$, $\tilde{\mathbf{H}} = \mathbf{\Theta}\mathbf{H}_t$ је еквивалентна матрица преноса која у себи обједињује матрицу $\mathbf{H}$ и матрицу кашњења $\mathbf{\Theta}$. Вектор $\tilde{\mathbf{N}}_t$, $\tilde{\mathbf{N}}_t = \mathbf{\Theta}\mathbf{N}_t$ је еквивалентни шум микрофона. Прва колона матрице преноса $\tilde{\mathbf{H}}$, вектор $\tilde{\mathbf{h}}_0$, описује пренос од одбраног говорника до сваког микрофона. Вектор $\tilde{\mathbf{h}}_0$ можемо декомпоновати на пренос директног таласа $\mathbf{h}_{d0}$ и на пренос рефлектованих таласа $\mathbf{h}_{r0}$ према релацији

$$\widetilde{\mathbf{h}}_0 = \mathbf{h}_{d0} + \mathbf{h}_{r0} \ , \qquad (3)$$

Вектор преноса директног таласа $\mathbf{h}_{d0}$ има све јединице. Пренос рефлектованих таласа $\mathbf{h}_{r0}$ обједињује у себи релативно кашњење рефлексија у односу на директан талас за фиксну вредност $\tau_r$ и филтер минималне фазе.

Директни талас детектујемо тестирањем следећих двеју хипотеза H0 и H1:

H0: Присутан је само директан талас одабраног говорника и шум, што описујемо моделом $\mathbf{M}_0$ са

$$\mathbf{X}_t = \mathbf{h}_{d0} s_{0t} + \widetilde{\mathbf{N}}_t \qquad (4)$$

H1: Поред директног таласа постоје рефлексије и сметње. Овај сложенији модел $\mathbf{M}_1$ описујемо релацијом

$$\mathbf{X}_t = \widetilde{\mathbf{H}}_t \mathbf{S}_t + \widetilde{\mathbf{N}}_t \qquad (5)$$

## 3. ВЕРОДОСТОЈНОСТ ХИПОТЕЗЕ H0

Модел $\mathbf{M}_0$ (4), можемо интерпретирати као регресиони модел првог реда са следећим значењима:

$\mathbf{X}_t$ - Вектор од $n$ опсервација,

$\mathbf{h}_{d0}$ - Познат вектор димензије $n$, $\mathbf{h}_{d0} = [1,...,1]^*$ ,

$s_{0t}$ - Непознати параметар бројно једнак корисном сигналу у тренутку $t$.

Користећи (4), средњеквадратну процену сигнала $s_{0t}$ у тренутку $t$ добијамо релацијом

$$\hat{s}_{0t} = \frac{1}{\mathbf{h}_{d0}^* \mathbf{h}_{d0}} \mathbf{h}_{d0}^* \mathbf{X}_t = \frac{1}{n}[1,...,1]\mathbf{X}_t \ , \qquad (6)$$

где је са $^*$ означено коњуговано-комплексно транспоновање матрице (вектора). Процена вектора грешке у тренутку $t$ једнака је

$$\mathbf{E}_t = \mathbf{X}_t - \mathbf{h}_{d0}\hat{s}_{0t} \ . \qquad (7)$$

Под претпоставком да шум на микрофонима има Гаусову расподелу са варијансом σ, логаритам веродостојности да је сигнал $\mathbf{X}_t$ генерисан моделом $\mathbf{M}_0$ у тренутку $t$, сагласно [5] једнак је

$$L_t = \log\left(p(\mathbf{X}_t \mid \hat{s}_{0t})\right) = n \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{\left(\mathbf{E}_t^* \mathbf{E}_t\right)}{2\sigma^2} \ . \qquad (8)$$

## 4. ДЕТЕКЦИЈА ПРЕКО ЈЕДНОГ МОДЕЛА

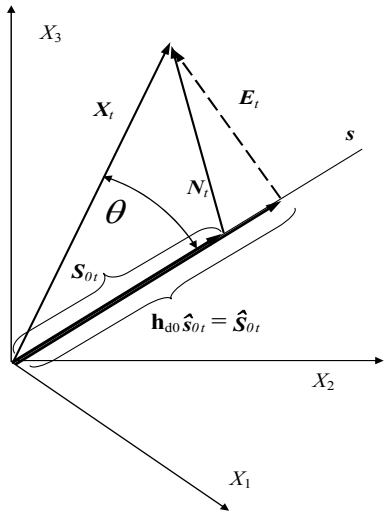
Приступ детекције преко једног модел [5] користи се када нам је познат само један модел (пре промене), док нам други модел (након промене) није познат. У том случају се детекција промене модела $\mathbf{M}_0$ у модел $\mathbf{M}_1$ врши надзирањем функције веродостојности модела $\mathbf{M}_0$. Уколико веродостојност постане мања од усвојеног прага одлуке λ, констатује се да је наступила промена и да више не важи модел $\mathbf{M}_0$. Сагласно овоме, правило одлучивања описује се релацијом

$$L_t \underset{H_0}{\overset{H_1}{\lessgtr}} \lambda \qquad (9)$$

Тешкоћа у примени релације (8) је што унапред не знамо варијансу шума σ. Овај проблем се може превазићи геометријским приступом.

## 5. ГЕОМЕТРИЈСКИ ПРИСТУП

Због очигледнијег тродимензионалног графичког приказа, геометријски приступ ћемо описати на примеру тро-микрофонског низа (слика 2). Вектор директног таласа корисног сигнала $\boldsymbol{S}_{0t}$ има све три компоненте бројно једнаке, те лежи на правој **s**.




Сл. 2. *Просторни приказ вектора тренутних сигнала.*

Вектор $\hat{\boldsymbol{S}}_{0t}$ је пројекција вектора мерења $\mathbf{X}_t$ на осу **s,** те представља средњеквадратну процену $\boldsymbol{S}_{0t}$ када важи модел $\mathbf{M}_0$. Дефинишимо функцију одлуке $f(\mathbf{X}_t)$ као реципрочну вредност квадрата синуса угла $\theta$ релацијом

$$f(\mathbf{X}_t) = \sec^2(\theta) = \frac{\left\|\mathbf{X}_t\right\|^2}{\left\|\mathbf{E}_t\right\|^2} \ , \qquad (10)$$

где је са $\|\,.\,\|$ означена Еуклидска норма вектора. Обе функције $f(\mathbf{X}_t)$ и $L_t$ су строго опадајуће у односу на норму вектора грешке $\|\,\mathbf{E}_t\,\|^2$. Стога, уместо логартима веродостојности $L_t$, одлука се може донети на основу функције $f(\mathbf{X}_t)$

$$f(\mathbf{X}_t) \underset{H_0}{\overset{H_1}{\lessgtr}} \lambda \qquad (11)$$

Уопштавањем релације (8) на случај када нам је познато $N$ узастопних пакета одбирака сигнала, веродостојност да су они генерисани моделом $\mathbf{M}_0$ једнака је,

$$L_{t:t+N-1} = Nn \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \sum_{k=t}^{t+N-1} \frac{\left\|\mathbf{E}_k\right\|^2}{2\sigma^2} \ . \qquad (12)$$

Аналогно релацији (12), можемо дефинисати геометријску функцију одлуке за $N$ узастопних пакета одбирака сигнала релацијом

$$f(\mathbf{X}_t,...,\mathbf{X}_{t+N-1}) = \frac{\sum_{k=t}^{t+N-1}\left\|\mathbf{X}_k\right\|^2}{\sum_{k=t}^{t+N-1}\left\|\mathbf{E}_k\right\|^2} \ , \qquad (13)$$

са правилом одлучивања

$$f(\mathbf{X}_t,...,\mathbf{X}_{t+N-1}) \underset{<}{\overset{H_1}{\gtrless}} C_\lambda, \quad (14)$$
$$H_0$$

где је $C_\lambda$ је усвојени праг одлуке.

## 6. ОПТИМИЗАЦИЈА АЛГОРИТМА

На говорном сигналу предложени алгоритам детекције показује одређене слабости из следећег разлога. После паузе у говору тренутна снага говорног сигнала обично расте постепено, без наглог скока снаге. Последица овога је прожимање директног таласа и рефлексија. Ово отежава детекцију. Да би се алгоритам учинио робуснијим у односу на поменуте услове детекције, предлажу се два побољшања.

Метод пондерисаних квадрата. Идеја је да се у рачунању квадрата грешака релацијом (13) истакну они делови сигнала са већом снагом и повољнијим односом сигнал/шум. Сагласно овоме релацију (13) треба заменити следећом

$$f_p(\mathbf{X}_t,...,\mathbf{X}_{t+N-1}) = \frac{\sum_{k=t}^{t+N-1} \alpha_k \left\|\mathbf{X}_k\right\|^2}{\sum_{k=t}^{t+N-1} \alpha_k \left\|\mathbf{E}_k\right\|^2}, \quad (15)$$

где су $\alpha_k$ пондеришући коефицијенти које израчунавамо са

$$\alpha_k = \frac{1}{p_{\max}} p_w(\mathbf{X}_k) \le 1,\ p_w(\mathbf{X}_k) = \left\|\mathbf{X}_k\right\|^2, \quad (16)$$
$$p_{\max} = \max_k \{p_w(\mathbf{X}_k)\},\quad k = 1, N\ .$$

Метод предфилтрирања. Све спектралне компоненте сигнала не доприносе подједнако успешности детекције. Познато је да је највећа енергија говорног сигнала сконцентрисана у опсегу испод 1000Hz, где је и најповољнији однос сигнал/шум. Стога је корисно улазне микрофонске сигнале филтрирати линеарним филтрима пропусницима учестаности до 1000Hz, чиме се елиминише утицај компонената које због ниског односа сигна/шум неповољно утичу на тачност детекције.

## 7. ПРИМЕНА У ML АДАПТИВНОМ ФИЛТРИРАЊУ

Критеријум максималне веродостојности у потискивању амбијенталних сметњи микрофонским низом има погодност да се његовом применом максимално потискују присутне сметње, а истовремено истиче користан сигнал [6]. Са друге стране, примена овог критеријума захтева познавање двеју коваријационих матрица. Првом, сигналном коваријационом матрицом описује се одзив просторије на користан сигнал, док се другом, коваријационом матрицом сметњи, описује одзив просторије на присутне сметње [6]. У условима једновременог деловања корисног сигнала и сметњи ове матрице није могуће естимирати из доступних мерења микрофонских сигнала.

Естимација коваријационих матрица сигнала и сметњи у случају говорних сигнала ипак је могућа, будући да у просеку 20 процената времена у говору припада паузама. Због великог процента пауза, постоје временски интервали када су присутне само сметње и интервали када је присутан само користан сигнал. Циљ је да се применом одређених алгоритама издвоје интервали са паузама у говору, и интервали са јаким односом користан сигнал/шум и да се на њима естимирају потребне коваријационе матрице. Детекција пауза у корисном сигналу може се успешно реализовати алгоритмом изложеним у [2], док се интервали са говором могу детектовати предложеним поступком детекције директног таласа. Након детекције директног таласа који траје врло кратко, следи дужи интервал који садржи одзив (реверберацију) просторије на користан сигнал, који се може искористити за естимацију коваријационе матрице сигнала. На тако естимираним коваријационим матрицама сигнала и сметњи примењује се поступак естимације оптималних тежинских коефицијената описан у [6].

## 8. ЕКСПЕРИМЕНТАЛНИ РЕЗУЛТАТИ

У циљу оцене успешности детекције предложеног алгоритма и његове примењивисти на ML алгоритам симулирана је соба са реверберацијом $T_{60}$=270*ms* методом ликова у огледалу [7]. Симулирана су два говорника од којих је први $s_1$ био циљ, док је други $s_2$ био сметња [2]. Микрофонски низ се састојао од 8 микрофона са међумикрофонским одстојањем од 6*cm*. Учестаност одабирања аудио сигнала је износила 10KHz, а њихово трајење је било 10s. Да би добро моделирале преносне функције таласа у условима релативно јаке реверберације, коришћен је DFT са 4096 тачака.

Резултат детекције директног таласа је приказан на слици 3 где је на диграмима приказано а) временски облик циљног говорника, b) временски облик говоника-сметње, c) временски облик њихове суперпозиције на микрофону 1, d) временски облик функције одлуке $f(\mathbf{X}_t,...,\mathbf{X}_{t+N-1})$. Локални максимуми функције $f(\mathbf{X}_t,...,\mathbf{X}_{t+N-1})$ указују на делове сигнала где је доминантан директан талас циљног говорника. Већи локални максимуми јављају се тамо где је однос сигнал/сметње повољнији.

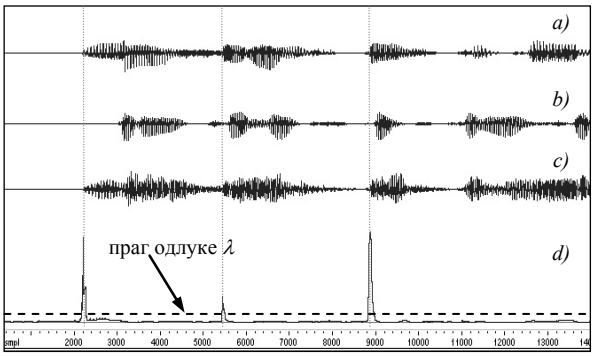
Након детекције директног таласа следи интервал где су присутни и директан талас и његове рефлексије. Трајање овог интервала предложеним поступком није могуће одредити. Стога се као хеуристичко решење предлаже да се након детекције директног таласа, за потребе естимације коваријационе матрице сигнала издвоји временски интервал фиксне дужине. Експериментално је констатовано да је за добру процену коваријационе матрице, довољно да тај интервал буде дужине 100ms. Коваријациона матрица сметњи се естимира поступком предложеним у [2]. Обе естимиране коваријационе матрице се затим користе у ML алгоритму [6] за одређивање оптималних тежинских коефицијената **W**.

Експериментално су упоређивани следећи алгоритми:

1) CBF (Conventional Beamformer), потискивач са фиксним коефицијентима.
2) Уобичајени GSC (Generalized Sidelobe Canceller) потискивач са пуним степеном адаптације.
3) GSC алгоритам код кога је адаптација тежинских коефицијената вршена на ручно одређеним паузама у корисном сигналу као у [3].

4) GSC алгоритам код кога је адаптација тежинских коефицијената вршена под идеалним условима када користан сигнал није уопште био присутан као у [3].
5) ML алгоритам са предложеним поступком процене коваријационих матрица сигнала и сметњи.
6) ML алгоритам код кога су коваријационе матрице сигнала и сметњи процењене под идеалним условима. Коваријациона матрица сигнала је процењивана када је само користан сигнал био присутан, а коваријациона матрица сметњи на сегментима где је само сметња била присутна.

Резултати експеримента приказани су у табели Т1 где су за поједине алгоритме дате кепстралне мере одстојања у односу на одзив просторије на побуду говорника $s_1$. Алгоритми су сортирани према растућем квалитету реконструкције корисног сигнала.




Сл.3. *а) циљни говорник $s_1$, b) говорник сметња $s_2$, c) суперпозиција $s_1$ и $s_2$ на микрофону 1, d) критеријумска функција* $f(\mathbf{X}_t,...,\mathbf{X}_{t+N-1})$.

Табела 1. Кепстралне мере одстојања за поједине алгоритме

| Алгоритам | Кепстрална мера |
|---|---|
| 1. CBF | 0.860 |
| 2. GSC-пун степен адаптације | 0.758 |
| 3. GSC-ручно одређене паузе | 0.607 |
| 4. GSC идеалан сценарио | 0.524 |
| 5. ML GSC са предложеним поступком детекције директног таласа | 0.496 |
| 6. ML GSC –идеалан сценарио | 0.453 |

## 5. ЗАКЉУЧАК

У раду је показано да се детекција директног таласа одабраног говорника у просторији са реверберацијом, може успешно реализовати применом функције одлуке базиране на средњеквадратној мери одступања од усвојеног модела директног таласа. Поступак је базиран на познавању модела генерисања директног таласа, а детектује се средњеквадратно отступање од усвојеног модела. Показано је да се поступак може успешно применити на естимацију коваријационе матрице сигнала за потребе ML алгоритма адаптивног формирања пријемног снопа микрофонских низова. Потенцијално подручје примене је у мерењима и анализи акустичких особина просторије.

**Захвалница:** Овај рад је подржан он Министарства за науку и технологију Републике Србије кроз пројекат OI-1784.

## ЛИТЕРАТУРА

[1] J. E. Greenberg, "Evaluation of an adaptive beamforming method for hearing aids," JASA, 1662-1676, (1992).
[2] Zoran M. Šarić, Slobodan T. Jovičić, "Adaptive microphone array based on pause detection", *Acoustics Research Letters Online (ARLO) 5(2)*, pp 68-74 April 2004.
[3] O. Hoshuyama, "A realtime robust adaptive microphone array controlled by an SNR estimate," Proc. ICASSP98, pp. 3605-3608.
[4] Zoran Šarić, Slobodan Jovičić, "Subband pause in speech signal detection using microphone array in room with reverberation", *Proceedings of the* SPECOM 2004, 20-22 October, 2004, St. Petersburg, pp 132-137.
[5] M. Basseville, Igor V. Nikiforov, "Detection of Abrupt Changes: Theory and Applications", Prentice Hall, April 1, 1993.
[6] Zoran Šarić, Slobodan T. Jovičić, Srbijanka Turajlić, 2nd International Conference on Fundamental and Applied Aspects of Speech and Language, 29 November – 1 December, 2004, Belgrade, pp. S9.5.
[7] Jont B. Allen, David A. Berkley, "Image method for efficiently simulating small-room acoustics", *J. Acoust. Soc. Amer.* Vol.65, no.4, pp 943-950, Apr.1979.

**Abstract -** The labeling of the speech segments that belongs to desired speaker or interferences is important for the analysis and processing of the microphone array signals recorded in the room with reverberation and cocktail party interferences. Speech segments labeling proposed in this paper is based on the direct wave detection. The idea is to detect time delay interval between direct wave and reflections of the walls. The algorithm tests two hypotheses: H0 – there is a direct wave segment, and H1 – there is a mixture of the direct wave, reflections and interferences. Detection algorithm uses least squares error of the assumed direct wave propagation model. The signals are processed in time domain that provides small time delay of the detection.

The proposed detection algorithm is experimentally verified by simulating the room with reverberation. The detection algorithm applied to the ML algorithm for microphone array weights estimation provides better quality of the restored speech signal compared to the ordinary used minimum variance estimation algorithms.

## SPEECH SEGMENTS ANALYSIS IN REVERBERANT ROOM AND COCKTAIL PARTY INTERFERENCES USING MICROPHONE ARRAY

Зоран Шарић, Слободан Јовичић