



DOI: 10.5862/JCSTCS.234.3

УДК 621.391.84

*С.В. Перелыгин, Л.Х. Нурмухамедов*

## ПОВЫШЕНИЕ КАЧЕСТВА ОБРАБОТКИ РЕЧЕВЫХ СИГНАЛОВ С ПОМОЩЬЮ МИКРОФОННОЙ АНТЕННОЙ РЕШЕТКИ

*S.V. Perelygin, L.H. Nurmukhamedov*

### IMPROVING THE QUALITY OF SPEECH SIGNAL PROCESSING VIA A MICROPHONE ARRAY

Смоделирован процесс обработки широкополосного акустического сигнала с помощью микрофонной решетки применительно к задаче пространственного разделения двух акустических сигналов при их одновременном воздействии на решетку. Использовано взвешенное суммирование в частотной области; сделаны пороговые ограничения на весовые коэффициенты микрофонной решетки с целью минимизации искажений полезного сигнала, прошедшего обработку. По результатам экспертных статистических испытаний определены оптимальные пороговые значения весовых коэффициентов.

АКУСТИЧЕСКАЯ АНТЕННА; ШИРОКОПОЛОСНЫЙ СИГНАЛ; ПОЛЕЗНЫЙ И МЕШАЮЩИЙ СИГНАЛЫ; ИСКАЖЕНИЕ СИГНАЛА; ВЗВЕШЕННОЕ СУММИРОВАНИЕ; ПОРОГОВОЕ ОГРАНИЧЕНИЕ; ЭКСПЕРТНОЕ ОЦЕНИВАНИЕ.

The topic of the article is about modeling broadband acoustic signal processing via a microphone array. The article solves the problem of tuning the microphone array to suppress the signal from the source at one angle and to receive the signal undistorted from the source at the other angle during the modeling. Weighted summation in the frequency domain is used; threshold limits are imposed on weight coefficients in order to minimize distortion of the useful signal during processing. The optimal threshold values of weighting coefficients have been determined according to the results of expert statistical tests. Setting these threshold values, we achieve the conditions when both the artifacts (hissing, whistling, ringing, clanging) and the wanted signal distortions (change in the voice timbre, muting consonants) are minimally noticeable.

ACOUSTIC ANTENNA; BROADBAND SIGNAL; WANTED AND INTERFERING SIGNALS; SIGNAL DISTORTION; WEIGHTED SUMMATION; THRESHOLD LIMIT; EXPERT EVALUATION.

Микрофонные решетки позволяют решать задачи пространственной фильтрации звука, что дает возможность принимать акустический сигнал выборочно, только по определенному одному или нескольким направлениям [1–5]. Настроенная решетка должна иметь минимальную чувствительность по направлению на мешающий источник, в то время как полезный сигнал (голос целевого диктора) должен быть принят и неискажен.

В радиотехнике известны способы решения данной задачи с помощью антенных решеток, принимающих радиосигналы [6–9]. Методы расчета антенных решеток

основаны на использовании моделей приема либо гармонического, либо узкополосного сигнала [6, 7]. Акустические сигналы являются широкополосными, что должно учитываться при решении задачи настройки микрофонной решетки.

#### Расчетные соотношения

Как показано в [10], уже при наличии двух идентичных ненаправленных микрофонов (случай вырожденной микрофонной решетки) можно осуществить пространственное разделение двух широкополосных сигналов, источники которых образуют углы  $\varphi'$  и  $\varphi''$  относительно нормали к решетке (рис. 1).

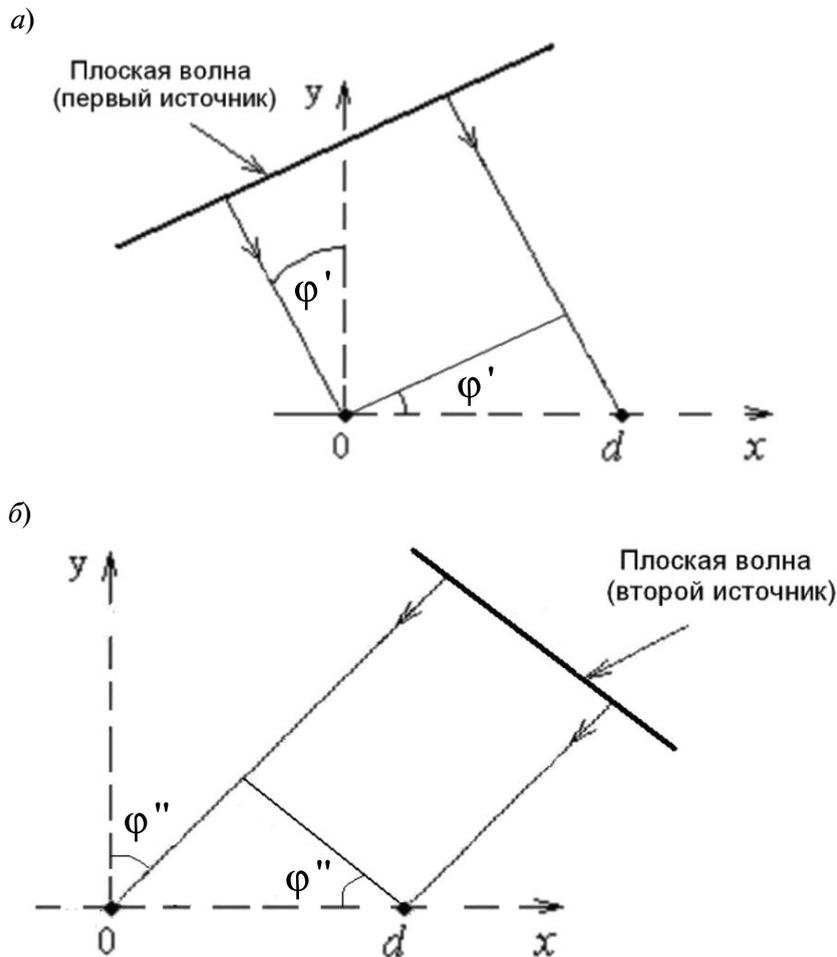


Рис. 1. Схема, поясняющая принцип работы решетки:  
а – звуковая волна от первого источника; б – звуковая волна от второго источника

Для решения задачи применяется взвешенное суммирование принятых сигналов в частотной области и оговаривается, что воздействующие на решетку сигналы занимают одинаковый диапазон частот и имеют плоский волновой фронт. Сигнал на выходе настроенной решетки находится из соотношения [6, 10]:

$$\dot{S}(j\omega) = w_1(j\omega) \cdot \dot{S}_1(j\omega) + w_2(j\omega) \cdot \dot{S}_2(j\omega), \quad (1)$$

где  $\dot{S}(j\omega)$ ,  $\dot{S}_1(j\omega)$ ,  $\dot{S}_2(j\omega)$  – спектральные плотности, соответственно, выходного сигнала, сигнала на выходе первого микрофона, сигнала на выходе второго микрофона;  $w_1(j\omega)$ ,  $w_2(j\omega)$  – частотно-зависимые весовые коэффициенты, рассчитываемые по следующим формулам [10]:

$$w_1(j\omega) = \frac{1}{1 - \exp\left(j\omega \frac{d(\sin \varphi'' - \sin \varphi')}{c}\right)}; \quad (2 \text{ а})$$

$$w_2(j\omega) = \frac{-\exp\left(j\omega \frac{d \sin \varphi''}{c}\right)}{1 - \exp\left(j\omega \frac{d(\sin \varphi'' - \sin \varphi')}{c}\right)}, \quad (2 \text{ б})$$

где  $c$  – скорость распространения звуковой волны;  $d$  – расстояние между микрофонами;  $\varphi'$  и  $\varphi''$  – углы прихода сигналов от первого и второго источников, соответственно;  $\omega$  – циклическая частота;  $j$  – мнимая единица.

### Моделирование

В ходе моделирования с использованием среды Matlab был имитирован процесс

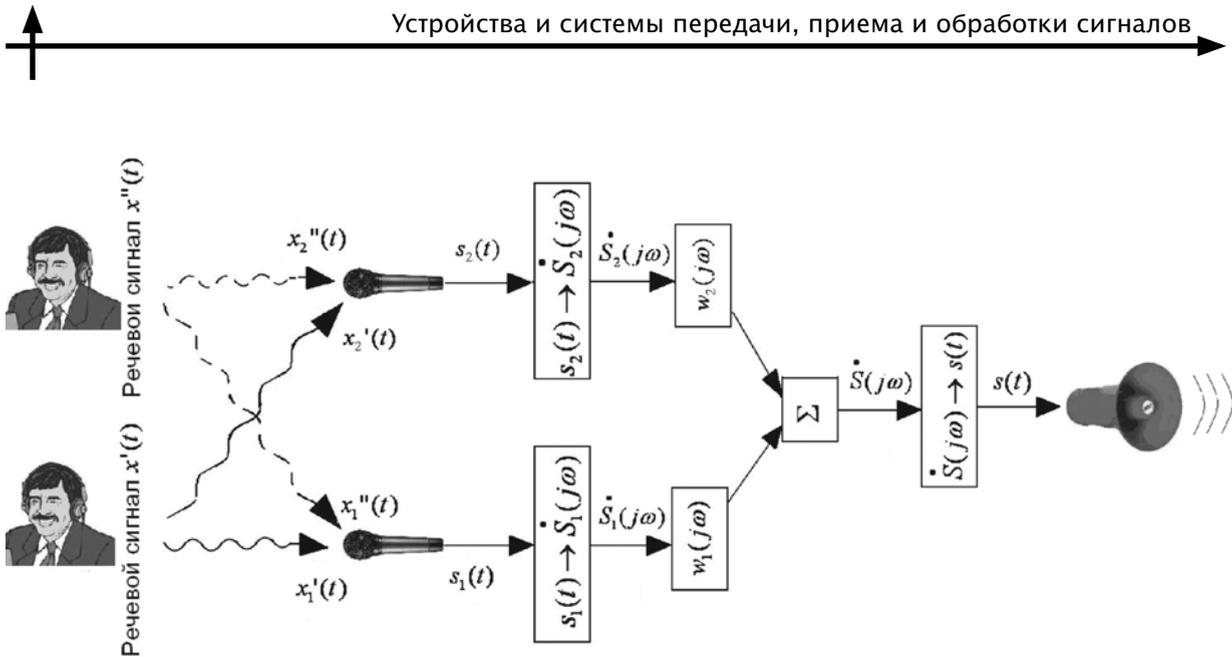


Рис. 2. Схема звукового тракта при проведении моделирования

прохождения сигналов через микрофонную решетку, состоящую из двух идентичных ненаправленных микрофонов. На рис. 2 схематично показан звуковой тракт, использованный при проведении моделирования.

В качестве исходного материала для моделирования применялись два речевых сигнала  $x'(t)$  и  $x''(t)$ , создаваемые одним человеком. Тем самым были выполнены условия, когда оба источника излучают акустические широкополосные сигналы в одинаковом диапазоне частот.

На подготовительном этапе произвольно задавались углы  $\varphi'$  и  $\varphi''$ . Формировалась группа сигналов:  $x'(t)$  и  $x''(t)$  — точная копия двух исходных сигналов,  $s_1(t)$  — их сумма (звуковая смесь на выходе первого микрофона),  $x_2'(t) = x_1'(t - \tau')$  и  $x_2''(t) = x_1''(t - \tau'')$  — сигналы, сдвинутые по времени относительно исходных сигналов на величины  $\tau'$  и  $\tau''$ , определяемые углами  $\varphi'$  и  $\varphi''$ ,  $s_2(t)$  — их сумма (звуковая смесь на выходе второго микрофона). Таким способом имитировался процесс прохождения звукового сигнала через первый и второй микрофоны.

На этапе обработки вычислялись весовые коэффициенты по формулам (2 а) и (2 б). Методом выборок с помощью дискретного преобразования Фурье (ДПФ) осуществлялся переход в частотную область

$s_1(t) \rightarrow \dot{S}_1(j\omega)$ ,  $s_2(t) \rightarrow \dot{S}_2(j\omega)$ . Использовались ДПФ со следующими параметрами: частота дискретизации 48 кГц, количество отсчетов 2048, временной период 42,7 мс, окно Ханна. Для спектра каждой выборки производилось взвешенное суммирование в соответствии с (1). В завершение данного этапа осуществлялся обратный переход из частотной области во временную  $\dot{S}(j\omega) \rightarrow s(t)$ , формируя результирующий сигнал из набора его выборок.

На этапе субъективного оценивания результатов выполнялось прослушивание и сравнение между собой сигналов  $s(t)$  и  $x_1'(t)$ .

### Результаты моделирования

Поскольку решетка была настроена на заранее известные углы  $\varphi'$  и  $\varphi''$ , то, как и ожидалось, в отклике решетки отсутствовал мешающий сигнал. Одновременно были выявлены искажения полезного сигнала в виде шипения и свиста.

Рассмотрим причины появления таких искажений. В процессе моделирования в подаваемом на решетку сигнале неизбежно присутствуют шумы помещения, собственные шумы микрофонов и шумы квантования. Можно предположить, что появление искажений возникает на тех частотах, для которых модули весовых коэффициентов имеют большие значения, что приводит к

подчеркиванию шумовых составляющих в сигнале.

При работе с широкополосными сигналами возникают критические случаи для набора частот, для которых знаменатели в (2 а) и (2 б) обращаются в нуль, а абсолютные значения весовых коэффициентов обращаются в бесконечность. Из (2 а) и (2 б) следует, что такие частоты «всплеска» равны:

$$f = \frac{c}{d |(\sin \varphi'' - \sin \varphi')|} k, k = 0, 1, 2 \dots \quad (3)$$

Для борьбы с отмеченными артефактами предлагается использовать метод порогового оценивания, т. е. ограничить сверху значение модуля весового коэффициента и субъективно оценить произошедшие изменения в обработанном сигнале. Если при расчете по формулам (2 а) и (2 б) весовой коэффициент по модулю получается выше порогового значения, то ему присваивается значение, равное пороговому.

Для проверки выдвинутой гипотезы был проведен модельный эксперимент по определению максимального порогового значения, при котором еще не слышны вносимые искажения звука (свист и шипение). Эксперимент показал правильность выдвинутой гипотезы и существование

порогового значения весового коэффициента, удовлетворяющего требованию отсутствия искажения звука. Одновременно оказалось, что если пороговое значение выбрать слишком маленьким, то в самом речевом сигнале возникают искажения, проявляющиеся в ухудшении слышимости согласных звуков.

При проведении моделирования был выбран диапазон поиска порогового значения от 50 до 0,5. По результатам моделирования было выбрано значение, лежащее внутри диапазона и равное 1,5. Выбор определялся отсутствием шипения и свиста с одной стороны и субъективным оцениванием слышимости согласных звуков с другой стороны. Также на основании проведенного моделирования было принято решение разделить диапазон слышимых частот на две области и использовать два пороговых значения весового коэффициента. Нижняя область охватывает частоты, лежащие в окрестности частоты первого «всплеска» ( $k = 0$ ), в верхнюю область попадают остальные критические частоты. Граница областей (т. е. значение частоты, разделяющей области) численно равна половине значения частоты второго «всплеска» ( $k = 1$ ). Как следует из (3):

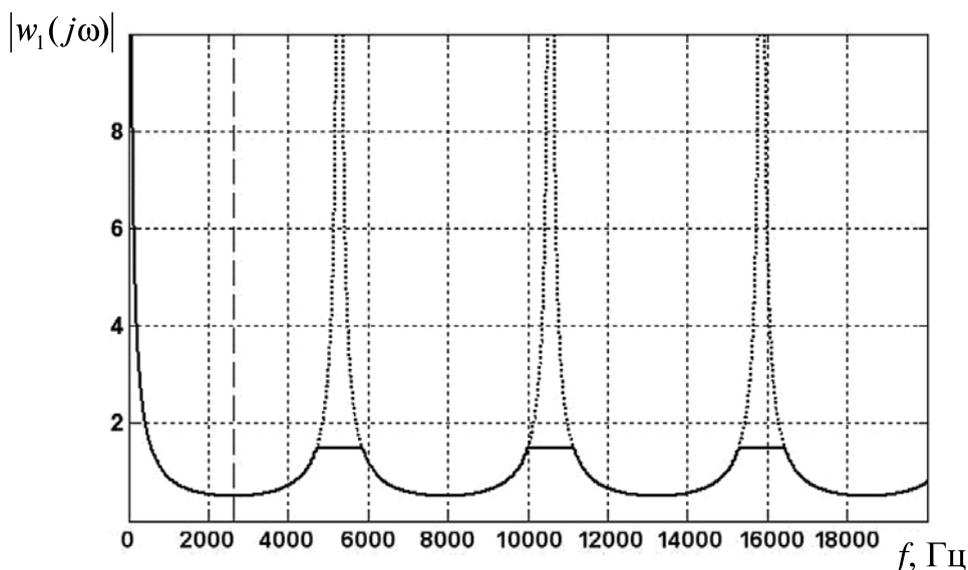


Рис. 3. Значения модуля весового коэффициента до (.....) и после (—) процедуры порогового ограничения

$$f_{\text{гран}} = \frac{c}{d |(\sin \varphi'' - \sin \varphi')|} \cdot 1/2. \quad (4)$$

Пороговое значение весового коэффициента в нижней области должно быть выше, чем в верхней. Выбор значения весового коэффициента в нижней области определялся с одной стороны отсутствием тембральных искажений голоса, а с другой стороны — отсутствием перегрузки звукового сигнала на низких частотах, которая приводила к снижению общего уровня сигнала при его прохождении через тракт. По результатам моделирования было выбрано пороговое значение, равное 100.

На рис. 3 представлены графики абсолютных значений весовых коэффициентов до и после процедуры порогового ограничения для углов  $\varphi' = -40^\circ$ ,  $\varphi'' = 40^\circ$ . Граница областей ( $f_{\text{гран}} = 2645$  Гц) показана пунктиром.

По аналогии были проведены модельные эксперименты для произвольных значений углов  $\varphi'$  и  $\varphi''$  от случая максимально возможного разнесения источников по на-

правлениям ( $\varphi' = -90^\circ$ ,  $\varphi'' = 90^\circ$ ) до случая их близкого расположения ( $\varphi' = -5^\circ$ ,  $\varphi'' = 5^\circ$ ). При этом использовались два постоянных пороговых значения весовых коэффициентов (100 для первой области, 1,5 для второй), изменялась только граница областей в соответствии с (4).

Результаты моделирования, полученные с использованием двух пороговых значений коэффициентов при точной настройке на углы  $\varphi'$  и  $\varphi''$ , показывают следующее:

мешающий речевой сигнал подавляется полностью;

вносимых искажений (шипения, свиста) не наблюдается;

слышимость согласных звуков для полезного сигнала не ухудшается;

полезный речевой сигнал иногда претерпевает тембральные искажения, не вызывающие раздражения у слушателя; в большинстве случаев тембральные искажения незаметны.

#### Экспертные испытания

Для окончательного установления опти-

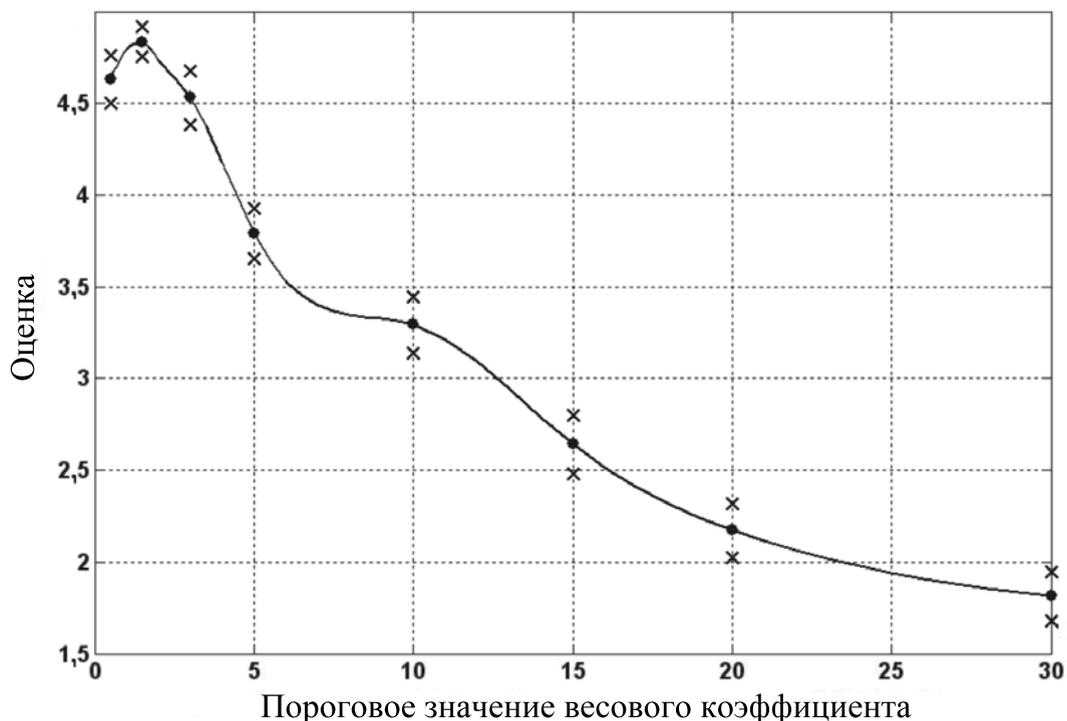


Рис. 4. Зависимость оценки искажений от порогового значения весового коэффициента в мелком масштабе

мального значения порога весового коэффициента в верхней области частот были проведены экспертные статистические испытания. В качестве исходного материала использовалась аудиозапись сигналов, полученных с выхода двух имитированных микрофонов. Эти сигналы представляли собой смесь двух голосов дикторов, находившихся под разными точно известными углами к решетке  $\varphi' = -40^\circ$ ,  $\varphi'' = 40^\circ$  и говоривших одновременно (см. рис. 1). В ходе каждого испытания задавалось пороговое значение весовых коэффициентов, и производился их расчет с учетом порогового ограничения. Далее осуществлялась обработка (разделение) сигналов путем взвешенного суммирования их спектров, и формировался результирующий сигнал, готовый для субъективной оценки. Затем задавалось новое пороговое значение, и формировался новый результирующий сигнал.

Полученная совокупность сигналов прослушивалась по очереди через головные телефоны и субъективно оценивалась. Для участия в экспертных испытаниях были привлечены преподаватели и студенты Санкт-Петербургского государственного института кино и телевидения (СПбГИКиТ) в общем количестве 92 человека. Перед каждым из экспертов была поставлена задача интегральной оценки заметности двух видов искажений: артефактов, проявляющихся при высоком значении порога в виде шипения, свиста и дребезга, и ухудшения качества самого сигнала, проявляющегося при низком значении порога в виде тембральных искажений и ухудшения слышимости согласных звуков. Для оценки заметности искажений, в соответствии с рекомендацией ITU-R BS.1284-1 [11], использовалась 5-балльная шкала со следующими градациями: 5 – искажения незаметны; 4 – искажения заметны, но не мешают; 3 – искажения немного мешают; 2 – искажения мешают и раздражают; 1 – искажения сильно мешают. Оценки, выставленные экспертами, далее подвергались статистической обработке.

Усредненные результаты субъективных оценок искажений полезного сигнала и доверительные интервалы представлены в виде графика на рис. 4.

Как следует из рисунка, оптимальное пороговое значение, являющееся компромиссом между искажениями первого и второго вида, равно 1,5. Таким образом, для минимизации искажений, возникающих в сигнале при обработке микрофонной решеткой, рекомендуется выбирать пороговое значение весовых коэффициентов, равное 1,5.

В статье рассмотрен метод, позволяющий адаптировать способы построения фазированной антенной решетки, используемой в радио- и гидролокации для узкополосного сигнала, применительно к широкополосному речевому сигналу. Показано, что прямая адаптация метода приводит к искажению речевого сигнала.

Для минимизации искажений проведены модельные эксперименты, по результатам которых даны рекомендации по выбору весовых коэффициентов в частотной области, используемых для получения взвешенного сигнала на выходе микрофонной решетки.

Эти рекомендации сводятся к определению границы между двумя областями слышимых частот, которым соответствуют два пороговых значения весовых коэффициентов, и к установлению пороговых значений весовых коэффициентов. Эти пороговые значения должны одновременно удовлетворять требованиям отсутствия шипения и свиста в сигнале, отсутствия ухудшения слышимости согласных звуков и отсутствия тембральных искажений.

В результате экспертных статистических испытаний получена зависимость оценки искажений сигнала от порогового значения весового коэффициента, позволяющая установить оптимальное пороговое значение, равное 1,5. При данном значении минимально заметны как вносимые искажения (шипение, свист), так и искажения самого сигнала (изменение тембра голоса, приглушение согласных звуков).

## СПИСОК ЛИТЕРАТУРЫ

1. **Столбов М.Б.** Применение микрофонных решеток для дистанционного сбора речевой информации // Научно-технический вестник информационных технологий, механики и оптики. 2015. Т. 15. № 4. С. 661–675.
2. **Brandstein M., Ward D.** (Eds.). *Microphone Arrays*. Heidelberg, Germany: Springer-Verlag, 2001.
3. **Huang Y.** *Microphone array signal processing* // Springer topics in signal processing. 2008.
4. **Tashev I.** *Sound Capture and Processing: Practical Approaches*. Wiley, 2009.
5. **Benesty J., Chen J.** *Study and Design of Differential Microphone Arrays* // Springer topics in signal processing. 2013.
6. **Монзинго Р.А., Миллер Т.У.** *Адаптивные антенные решетки: Введение в теорию*. Пер. с англ. М.: Радио и связь, 1986. 448 с.
7. **Бененсон Л.С., Журавлев В.А., Попов С.В., Постнов Г.А.** *Антенные решетки*. М.: Сов. радио, 1966. 368 с.
8. **Журавлев А.К., Лукошкин А.П., Поддубный С.С.** *Обработка сигналов в адаптивных антенных решетках: Монография*. Л.: Изд-во ЛГУ, 1983. 240 с.
9. **Пистолькорс А.А., Литвинов О.С.** *Введение в теорию адаптивных антенн*. М.: Наука, 1991. 200 с.
10. **Кривошейкин А.В., Перельгин С.В.** *Микрофонная решетка для реализации направленной акустической антенны* // Изв. вузов. Приборостроение. 2015. Т. 58, № 3. С. 221–225.
11. *Rec. ITU-R BS.1284-1. General methods for the subjective assessment of sound quality*. 1997–2003.

## REFERENCES

1. **Stolbov M.B.** *Primeneniye mikrofonnykh reshetok dlya distantsionnogo sbora rechevoy informatsii* [Application of microphone arrays for distant speech capture]. *Nauchno-tehnicheskiiy vestnik informatsionnykh tekhnologiy, mekhaniki i optiki* [Scientific and Technical Journal of Information Technologies, Mechanics and Optics]. 2015, Vol. 15, No. 4, Pp. 661–675. (rus)
2. **Brandstein M., Ward D.** (Eds.). *Microphone Arrays*. Heidelberg, Germany: Springer-Verlag, 2001.
3. **Huang Y.** *Microphone array signal processing*. Springer topics in signal processing, 2008.
4. **Tashev I.** *Sound Capture and Processing: Practical Approaches*. Wiley, 2009.
5. **Benesty J., Chen J.** *Study and Design of Differential Microphone Arrays*. Springer topics in signal processing, 2013.
6. **Monzingo R.A., Miller T.U.** *Adaptivnyye antennnyye reshetki: Vvedeniye v teoriyu* [Adaptive arrays: Introduction to the Theory]. Moscow: Radio i svyaz Publ., 1986, 448 p. (rus)
7. **Benenson L.S., Zhuravlev V.A., Popov S.V., Postnov G.A.** *Antennnyye reshetki* [Antenna arrays]. Moscow: Sov. Radio Publ., 1966, 368 p. (rus)
8. **Zhuravlev A.K., Lukoshkin A.P., Poddubnyy S.S.** *Obrabotka signalov v adaptivnykh antennnykh reshetkakh: Monografiya* [Signal processing in adaptive array antenna: Monograph]. Leningrad: LGU Publ., 1983, 240 p. (rus)
9. **Pistol Kors A.A., Litvinov O.S.** *Vvedeniye v teoriyu adaptivnykh antenn* [Introduction to adaptive antenna]. Moscow: Nauka Publ., 1991, 200 p. (rus)
10. **Krivosheykin A.V., Perelygin S.V.** *Mikrofonnaya reshetka dlya realizatsii napravlennoy akusticheskoy anteny* [Microphone array for directional acoustic antenna implementation]. *Izv. vuzov. Priborostryeniye* [Journal of Instrument Engineering], 2015, Vol. 58, No. 3, Pp. 221–225. (rus)
11. *Rec. ITU-R BS.1284-1. General methods for the subjective assessment of sound quality*, 1997–2003.

**ПЕРЕЛЫГИН Сергей Васильевич** – ассистент кафедры радиотехники и информационных технологий факультета технологий кино и телевидения Санкт-Петербургского государственного института кино и телевидения.

191119, Россия, Санкт-Петербург, ул. Правды, д. 13.

E-mail: sergey.perelygin@gmail.com

**PERELYGIN Sergey V.** St. Petersburg State University of Film and Television.

191119, Pravdy Str. 13, St. Petersburg, Russia.

E-mail: sergey.perelygin@gmail.com

**НУРМУХАМЕДОВ Леон Хасенович** — доцент кафедры радиотехники и информационных технологий факультета технологий кино и телевидения Санкт-Петербургского государственного института кино и телевидения, кандидат технических наук.

191119, Россия, Санкт-Петербург, ул. Правды, д. 13.

E-mail: nurmukh36@mail.ru

**NURMUKHAMEDOV Leon H.** *St. Petersburg State University of Film and Television.*

191119, Pravdy Str. 13, St. Petersburg, Russia.

E-mail: nurmukh36@mail.ru