

УДК 534.843.742

МЕТОД РАСШИРЕНИЯ ВОЗМОЖНОСТЕЙ СТЕРЕОФОНИИ, ОТНОСИТЕЛЬНО ПЕРЕМЕЩЕНИЯ СЛУШАТЕЛЯ В ПРОСТРАНСТВЕ

С. М. Порошин

Доктор технических наук, профессор*

В. В. Усик

Кандидат технических наук, доцент*

E-mail: sergeev_usik@ukr.net

И. С. Беликов

Аспирант*

E-mail: igorajon@yandex.ua

*Кафедра мультимедийных информационных технологий и систем

Национальный технический университет
«Харьковский политехнический институт»
ул. Фрунзе, 21, г. Харьков, Украина, 61002

Авторами у статті запропонована структурна схема системи автоматизованого управління переміщенням уявного випромінювання звуку у середовищі. Розглядається розширення можливостей відтворення акустичного стереосигналу відносно до переміщення людини у зоні прослуховування. Детектування людини відбувається за допомогою системи Microsoft Kinect, без додаткових розрізнявальних засобів, чи натільних датчиків. Положення уявного випромінювання звуку корегується шляхом змінення рівня сигналу та внесення часових затримок у канали акустичної системи

Ключові слова: OpenNI, стереофонія, Microsoft Kinect, уявне випромінювання звуку, Native Interaction, інтерактивні технології, мультимедійні технології

Авторами в статье предложена структурная схема системы автоматизированного управления перемещения кажущегося источника звука в пространстве. Рассматривается расширение возможностей воспроизведения акустического стереосигнала относительно перемещения человека в зоне прослушивания. Детектирование человека осуществляется при помощи системы Microsoft Kinect, без дополнительных опознавательных средств, или нательных датчиков. Положение кажущегося источника звука корректируется путем изменения уровня сигнала и внесения временных задержек в каналы акустической системы

Ключевые слова: OpenNI, стереофония, Microsoft Kinect, кажущийся источник звука, Native Interaction, интерактивные технологи, мультимедийные технологии

1. Введение

Интерактивное участие человека в управлении интерфейсом программ, а также других физических процессов, является одной из отличительных качеств современных мультимедийных технологий [1]. Использование технологии распознавания человеческого образа в пространстве и голосовое управление, позволяет постепенно избавляться от традиционных способов ввода информации [2]. На данный момент, среди мультимедийных акустических систем (АС) не используется активная корректировка воспроизведения кажущегося источника звука (КИЗ) в пространстве относительно перемещения слушателя. При отклонении человека от оси направленности громкоговорителя (ГГ), изменяется воображаемое расположение источников звука в помещении, где происходит озвучивание пространство. Это связано с изменением амплитудно-частотной характеристики (АС), при отклонении от оптимальной оси и изменением расстояния между ГГ акустической системы.

В данной работе представлена разработка структурной схемы аппаратно-программного комплекса по осуществлению управления перемещением КИЗ в пространстве на основе данных о местоположении

слушателя [3, 4]. Каждый ГГ обладает диаграммой направленности, и при отклонении от направления оси ГГ возникает изменение амплитудно-частотной характеристики воспроизведения сигнала. Внесение корректировок в амплитудно- и фазо-частотную характеристику акустической системы в реальном времени даёт возможность человеку постоянно находиться в оптимальной зоне прослушивания при перемещении в комнате, где он прослушивает мультимедийный контент [5].

Данные о местоположении головы слушателя и детектирование человека в пространстве осуществляется при помощи камеры Microsoft Kinect и соответствующей технологии OpenNI [4, 5]. Данная технология стала открытой для сторонних разработчиков в марте 2012 года, поэтому данная сфера является практически неисследованной и, на данный момент, широко развиваемой.

Применение интенсивностной и временной стереофония расширяет возможности воспроизведения звука акустической системы. Внесение задержки во времени в один из каналов акустической системы, также как и уменьшение уровня звукового сигнала, ведёт к смещению местонахождения кажущегося источника звука в противоположную сторону [6, 7].

Предложена экспериментальная методика определения центрального местоположения кажущегося источника звука акустической системы и дальнейшая калибровка его местоположения в пространстве относительно центра стереобазы.

2. Анализ литературных источников

Позиционирование КИЗ при озвучивании пространства рассчитывается исходя из свойств человеческого слуха. При расположении слушателя напротив центра стереобазы (точка О на рис. 1) АС, воспроизводимые сигналы с правого громкоговорителя ($Гр_1$) и левого громкоговорителя ($Гр_2$) не имеют различий по времени ($\Delta\tau=0$) и по уровню ($\Delta L=0$). При таком условии звучание обоих громкоговорителей сливается в единый звуковой образ, который соответствует изначально задуманному звукорежиссером звучанию. В процессе слияния образуется КИЗ [6].

Положение КИЗ на линии базы громкоговорителей зависит от временных ($\Delta\tau$) и интенсивностных (ΔL) различий между сигналами левого и правого каналов АС, достигающих ушей слушателя (рис. 1).

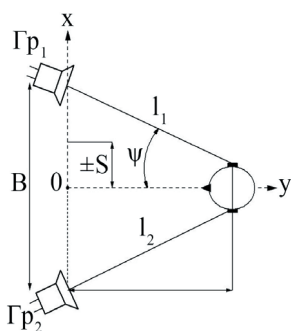


Рис. 1. Расположение слушателя относительно базы АС: В – база АС; О – центр стереобазы и место расположения камеры Kinect; $\pm S$ – диапазон смещения КИЗ от центрального положения; у – расстояние от центра базы АС до головы слушателя; l_1 – расстояние от головы слушателя до правого громкоговорителя $Гр_1$; l_2 – расстояние от головы слушателя до левого громкоговорителя $Гр_2$

Величина уровня сигнала ΔL для левого (L) и правого (R) каналов определяется по формуле (1).

$$\Delta L_{L,R} = 10 \lg \frac{l_1(x,y) D_2(\psi)}{l_2(x,y) D_1(\psi)}, \quad (1)$$

где l_1 – расстояние до первого громкоговорителя в координатах x,y ; l_2 – расстояние, соответственно, до второго громкоговорителя; D_1, D_2 – амплитудная характеристика громкоговорителя относительно угла отклонения ψ от оси АС [7, 8].

Расстояние от головы слушателя до $Гр_1$ и $Гр_2$ определяется следующим образом:

$$l_{1,2}(x,y) = \sqrt{\left(\frac{B}{2} \pm S(x)\right)^2 + y^2}, \quad (2)$$

где x, y – координаты слушателя в пространстве, В – размер установленной базы АС, $S(x)$ – смещение местоположения относительно центра стереобазы влево-вправо, y – расстояние отдаления слушателя от центра стереобазы.

Величина задержек $\Delta\tau$, вносимых в левый и правый канал соответственно, определяется взаимным местоположением громкоговорителей и слушателя в пространстве x,y формула (2).

$$\Delta\tau_{L,R} = \frac{l_1(x,y) - l_2(x,y)}{c}, \quad (3)$$

где c – скорость звука, l_1, l_2 – расстояние до соответственно первого и второго громкоговорителя в координатах x, y .

Если уменьшить громкость АС₁, то это будет воспринято слухом как перемещение КИЗ в сторону АС₂ и наоборот. Таким образом, варьируя громкость звучания левой и правой АС, можно вызывать и поддерживать иллюзию перемещения виртуального источника звука. Это явление называют интегральной локализацией (или локализацией суммы) [6–8].

При задержках одного из сигналов на время более 50 мс наличие запаздывающего сигнала ощущается как помеха в виде эха, хотя положение КИЗ остается неизменным.

Отсюда следует, что оригинальный сигнал, при одинаковом уровне с тем, в который внесена задержка, полностью подавляет (маскирует) последний. Повышая уровень запаздывающего сигнала, можно добиться результата, при котором оба источника звука будут восприниматься отдельно даже при запаздывании менее 50 мс [7].

Характер относительного перемещения КИЗ для малых баз ($B=0,8-1,8$ м) практически не зависит от расстояния y между слушателем и АС (рис. 1). Для относительно больших стереосистем с базой больше 2,8 м это утверждение справедливо при отдалении слушателя от центра базы на расстояние 2,8 м и более [6].

При $\Delta\tau < 5$ мс наблюдается неустойчивый режим: виртуальный источник звука перемещается из одной АС в другую, совпадая то с источником опережающего, то с источником задержанного сигнала. Если источники звука резко различаются по тембру, это может привести к отдельному ощущению двух звуковых объектов даже при равных уровнях интенсивности обоих сигналов. Малое время реверберации помещения (0,3-0,5 с) не влияет на локализацию звука [6–8].

3. Структурная схема активной системы коррективы местоположения кажущегося источника звука

Разработана структурная схема автоматизированной системы коррективы местоположения КИЗ в пространстве между каналами АС и головой слушателя (рис. 2).

Это необходимо для усиления эмоционального восприятия человека, при прослушивании аудио материала, а также при непосредственном интерактивном участии слушателя в создании акустической картины мультимедийного контента [9].

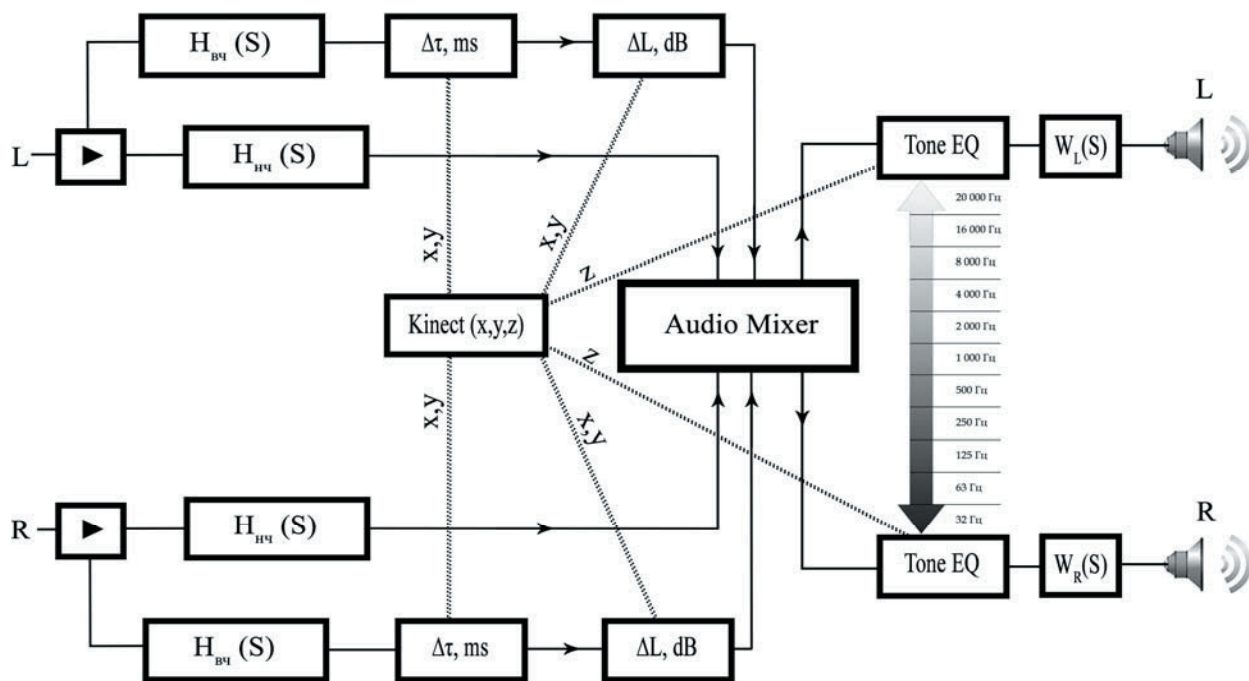


Рис. 2. Структурная схема автоматизированной коррекции перемещения КИЗ в пространстве

При маршрутизации стереосигнал разделяется на две составляющих по частотам. Бинауральный эффект восприятия звуков человеческом не позволяет точно определять источник КИЗ в пространстве при воспроизведении низких частот (НЧ).

Это обусловлено большой длиной волны звуковых колебаний, а также с особенностью строения человеческого слуха [7].

На частотах от 300 Гц и выше становится заметным сдвиг фаз звуковых волн, попадающих в правое и левое ухо. В данной работе предлагается вводить временные задержки $\Delta\tau_{x,y}$ и изменение интенсивности уровня сигнала $\Delta L_{x,y}$ в область частот от 300 Гц и выше.

На рис. 2 представлена маршрутизация акустического стереосигнала в системе по корректированию местоположения КИЗ относительно перемещения слушателя. С повышением частоты выше 1000 Гц становится заметным маскирование головой человека, то есть при повороте к источнику звука одним ухом, второе ухо находится в акустической тени головы и слабее ощущает звуковые колебания. Максимальное время запаздывания между ушами слушателя составляет 0,63 мс [10].

Наибольшая точность локализации достигается при восприятии сложных звуков и звуковых импульсов. При этом важным фактором является спектральный состав звуков. Наименьший ощутимый угол восприятия отклонения источника звуковых импульсов равен 3°. Эту величину следует считать разрешающей способностью слуха для фронтального направления. Точность локализации источников звука, расположенных слева и справа, значительно меньше и составляет примерно 12°. Для тылового направления эта величина около 6° [10].

Обработку акустического сигнала, в процессе маршрутизации (рис. 2) разделяется на несколько этапов.

4. Фильтрация сигнала на две составляющих

Индивидуально для левого и правого каналов, фильтрами нижних и верхних частот, с последующей корректировкой временных и интенсивностных характеристик в области рабочих частот (от 300 Гц и выше).

Фильтр нижних частот ($H_{НЧ}(S)$ на рис. 2) представляет собой устройство, пропускающее сигналы низких частот и подавляющее сигналы высоких частот [11]. Передаточная функция фильтра низких частот имеет вид:

$$H_{НЧ}(S) = \frac{K C \omega_c^2}{S^2 + B \omega_c S + C \omega_c^2}, \tag{4}$$

где ω_c – частота среза, B и C представляют собой нормированные коэффициенты, поскольку для $\omega_c = 1$ эта передаточная функция приводится к виду (5) при $n=2$.

$$H_{НЧ}(S) = \frac{K b_0}{s^n + b_{n-1} s^{n-1} + \dots + b_1 s + b_0}. \tag{5}$$

Фильтр верхних частот ($H_{ВЧ}(S)$ на рис. 2) представляет собой блок, пропускающий сигналы высоких частот и подавляющий сигналы низких частот. Передаточную функцию фильтра верхних частот с частотой среза ω_c можно получить из передаточной функции нормированного фильтра нижних частот (имеющего ω_c , равную 1 рад/с) с помощью замены переменной s на ω_c / s .

Следовательно, функция фильтров верхних частот Баттлерворта и Чебышева будет содержать следующие сомножители второго порядка:

$$H_{\text{ВЧ}}(S) = \frac{Ks^2}{s^2 + \left(\frac{B\omega_c}{C}\right)S + \frac{\omega_c^2}{C}}, \quad (6)$$

где ω_c – частота среза фильтра, а В и С представляют собой нормированные звена фильтра частот второго порядка [11–13].

Широкое применение нашли фильтры Бесселя, Баттерворта, Чебышева и эллиптические (Кауэра), названные так по виду полиномов передаточных функций, в разной степени аппроксимирующих амплитудно-частотную характеристику (АЧХ) идеальных фильтров и отличающиеся друг от друга крутизной наклона АЧХ в начале полосы задерживания и степенью колебаний переходного процесса при ступенчатом входном воздействии [12].

Ранее было описано, что дальнейшая обработка сигнала будет производиться в области частот от 300 Гц и выше, а отфильтрованная НЧ часть поступает без изменений на финальный аудио микшер (Audio Mixer на рис. 2).

Отфильтрованная область частот от 300 Гц и выше, далее в статье будет называться областью рабочей частоты [10].

В область рабочей частоты сигнала вносятся временные задержки, исходя из данных об удалении головы слушателя от центрального местоположения стереобазы.

На рис. 2 этот процесс указан в блоке Дт. Данные о местоположении слушателя в пространстве обозначаются как x, y, z.

Согласно рассчитанным данным, в аудио канал громкоговорителя, к которому человек расположен ближе, вносится временная задержка в мс. Величина Дт рассчитывается по формулам (2) и (3). Сигнал из противоположного громкоговорителя воспроизводится без задержки во времени. Это даёт возможность сместить КИЗ обратно в центр акустической базы, не нарушая целостности акустического восприятия материала [10].

Интенсивная стереофония (блок ΔL на рис. 2) позволяет скорректировать отклонение КИЗ путём усиления уровня сигнала канала, который подвергается маскированию, относительно местоположения головы слушателя. При приближении слушателя к одному из громкоговорителей, КИЗ смещается в сторону громкоговорителя, возле которого расположился слушатель.

Согласно формулам (1) и (3), применение интенсивной стереофонии позволяет скомпенсировать отклонение КИЗ путём ослабления уровня сигнала, находящегося ближе к слушателю [12–13].

Получение выходного сигнала выполняется путём микширования ранее отфильтрованной низкочастотной составляющей и диапазона частот, в которые вводились коррективы (Audio Mixer на рис. 2).

Возможность получения данных о местоположении головы слушателя в вертикальной плоскости (z на рис. 2), относительно установленных громкоговорителей, даёт возможность применения активной стереофонии в вертикальной плоскости.

В следующем разделе предложен концепт автоматизированной коррективы звучания АС в вертикальной плоскости.

5. Активная стереофония в вертикальной плоскости

Детектирование перемещения человека в вертикальной плоскости даёт возможность расширения стереофонии не только по отношению к удалению слушателя в пространстве относительно каналов громкоговорителей, но и внести корректирование АЧХ акустического сигнала, относительно движения в вертикальной плоскости. Данное преобразование КИЗ актуально в активных компьютерных играх, где человек принимает активное участие в управлении своим компьютерным персонажем при помощи перемещения своего тела.

Согласно теории известного звукорежиссера Боба Овсинки [14], панорамирование звуков в вертикальном пространстве относительно установленных громкоговорителей возможно при помощи корректировки АЧХ при отклонении от оси оптимальной направленности излучателей АС. Область нижних частот воспринимается человеческим сознанием в нижней части поля зрения, область верхних частот, соответственно вверх (рис. 3).

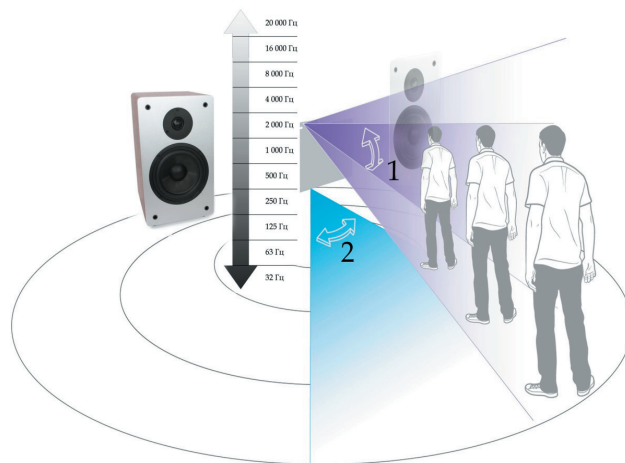


Рис. 3. Зависимость психофизического восприятия распространения частот в вертикальной плоскости относительно месторасположения головы слушателя: 1 – диапазон детектирования перемещения головы слушателя в вертикальной плоскости; 2 – диапазон детектирования перемещения слушателя в горизонтальной плоскости

При отклонении слушателя от оси громкоговорителя, изменяется АЧХ воспринимаемого сигнала. На рис. 4 показано АЧХ громкоговорителя при отклонении местоположения микрофона от перпендикулярного угла относительно динамиков громкоговорителя. АЧХ 1 – при расположении микрофона строго перпендикулярно относительно динамиков АС на высоте 1 м от уровня пола, АЧХ 2 соответствует отклонению в вертикальной плоскости до уровня 0,5 м от уровня пола в помещении.

На рис. 2 корректор АЧХ направленности громкоговорителей указан, как Tone EQ и является однополосным параметрическим эквалайзером. Центральную частоту фильтра предполагается смещать относительно вертикального перемещения головы слушателя в пространстве, область 1, рис. 3. Добротность фильтра мо-

жет быть настраиваемой в дальнейшем самостоятельно слушателем, исходя из его индивидуальных предпочтений. Вторым вариантом использования активной вертикальной стереофонии – это компенсация отклонения неравномерности АЧХ, как указано на рис. 4.

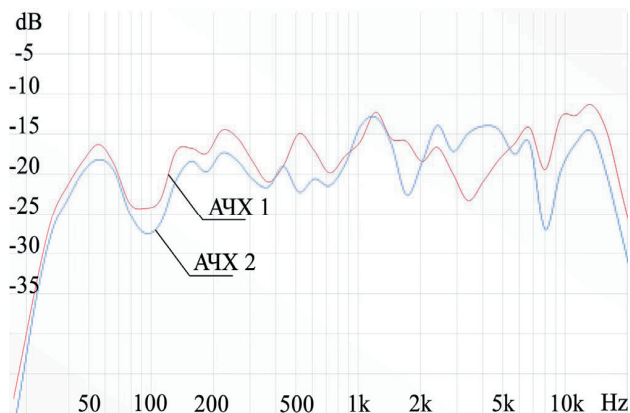


Рис. 4. Изменение неравномерности АЧХ при отклонении слушателя относительно громкоговорителя в вертикальной плоскости

Данная методика расширения стерео в вертикальной плоскости на данный момент мало изучена, в дальнейшем, авторами работы предполагается проведение дополнительных расчётов и экспериментальной проверки актуальности темы.

6. Методика экспериментальной проверки местоположения кажущегося источника звука

Авторами статьи были проведены экспериментальные оценки значений амплитуды сигнала, при использовании интенсивной стереофонии для корректировки КИЗ. В качестве измерителя использовался микрофон Zoom H1, который конструктивно являлся стереомикрофоном X-Y типа. Метод интенсивной стереофонии заключается в корректировке пространственного местоположения КИЗ путём увеличения (уменьшения) уровней сигнала левого и правого каналов, но у акустических систем разных моделей изначальное центральное местоположение КИЗ может отличаться от геометрического центра базы.

В данной работе авторами приводится методика определения центрального местоположения КИЗ и его последующая корректировка. Предлагается для левого и правого канала использовать синусоидальные сигналы разных частот, но лежащих в пределах одной октавы. Для левого канала АС используется тестовый сигнал с частотой 1000 Гц (на рис. 5 график АЧХ-L), для правого канала частота тестового сигнала составляет 1250 Гц (на рис. 5 график АЧХ-R). Выбор таких частот обосновывается нечётностью относительно гармонических искажений, которые являются кратными для исходных частот тестовых сигналов, таким образом, удаётся минимизировать влияние гармонических искажений левого канала на правый.

При совпадении величин амплитуды сигналов, поступающих с левого и правого каналов АС, можно сделать вывод о том, что выбранное местоположение

стереомикрофона не соответствует местоположению относительно центра КИЗ, даже если это совпадает с геометрическим центром базы АС.

На рис. 5 указана АЧХ сигнала, полученного стереомикрофоном, установленным напротив центра базы АС. Как видно из величин амплитуды сигнала на частоте 1250 Гц, интенсивность левого канала АС в правом канале микрофона превышает интенсивность правого канала АС в левом канале микрофона на 6 дБ. Это даёт смещение КИЗ в сторону левого громкоговорителя АС.

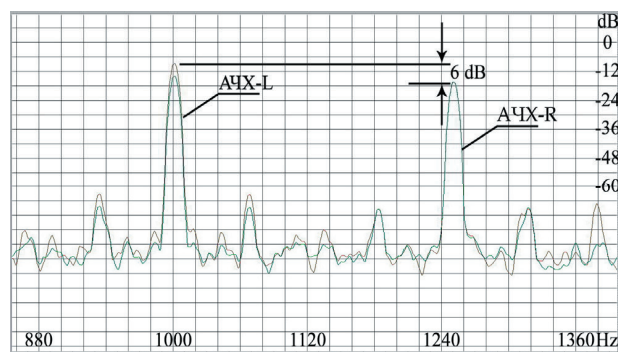


Рис. 5. АЧХ тестового стереосигнала при расположении микрофона относительно геометрического центра базы АС

Уменьшение интенсивности воспроизведения левого канала позволит откалибровать центральное местоположения КИЗ в пространстве относительно местоположения микрофона, которое соответствует местоположению головы слушателя.

На рис. 6 указана АЧХ тестового сигнала, полученного после компенсации уровня сигнала левого громкоговорителя. Уменьшение амплитуды сигнала левого канала АС на 6 дБ позволяет выровнять её уровень с амплитудой сигнала правого канала АС.

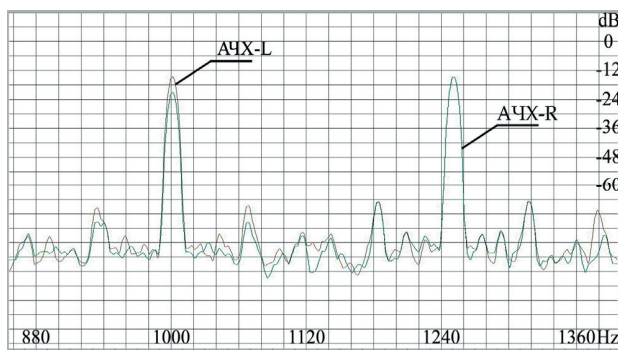


Рис. 6. АЧХ тестового сигнала при введении компенсации -6 дБ в левый канал тестового стереосигнала

Равнозначность амплитуд тестового сигнала на частотах 1000 и 1250 Гц, как для левого, так и для правого каналов, свидетельствует о верно подобранном местоположении стереомикрофона. Данное местоположение микрофона соответствует идеальному местоположению головы человека при прослушивании стереокомпозиций, центр КИЗ будет формироваться точно напротив головы слушателя.

7. Выводы

На данный момент, среди мультимедийных акустических систем не используется активная коррективка воспроизведения кажущегося источника звука в пространстве относительно перемещения слушателя. Использование технологии распознавания и детектирования человека в пространстве, позволяет использовать полученные данные для усиления интерактивности участия человека в формировании акустического сопровождения мультимедийного контента.

Предложена структурная схема автоматизированной системы коррективки местоположения кажущегося источника звука относительно перемещения слушателя в пространстве. Объединены разработки в области распознавания образа человека Microsoft OpenNI и обработки звукового контента.

Внесение задержки во времени в один из каналов акустической системы ведёт к смещению местонахождения кажущегося источника звука в противоположную сторону. Синхронизация данных о местоположении головы слушателя и величины необходимой задержки звукового сигнала между каналами акустической системы даёт возможность управления пере-

мещением КИЗ в реальном времени. Таким образом, слушатель становится интерактивным участником отображения мультимедийного контента.

Управление перемещением кажущегося источника звука осуществляется при использовании интенсивностной стереофонии и внесение корректировок в амплитудно-частотную характеристику акустической системы.

Так как уровень сигнала левого и правого канала в большинстве бытовых акустических аудио систем отличается друг от друга, для корректной оценки перемещения КИЗ необходима калибровка усилителей каналов в акустической системе.

Приведена методика определения центрального местоположения кажущегося источника звука и методика его коррективки относительно геометрического центра акустической базы в озвучиваемом пространстве.

Подтверждено экспериментально изменение местоположения кажущегося источника звука при использовании интенсивностной стереофонии.

Дальнейшим развитием является исследования математической модели автоматизированной системы управления акустическим трактом и сбор экспериментальных данных.

Литература

1. Порошин, С. Метод коррекции акустического сигнала в соответствии с перемещением зоны прослушивания [Текст] / С. Порошин, И. Беликов // Восточно-Европейский журнал передовых технологий. – 2013. – Т. 3, № 9(63). – С. 4–6. – Режим доступа: <http://journals.uran.ua/eejet/article/view/14840>
2. Kinect for Windows. Human Interface Guidelines v 1.7 [Text] / Microsoft Corporation, 2013. – 135 с.
3. Miles, R. Start Here! Learn Microsoft Kinect API [Text] / R. Miles; O'Reilly Media, Inc. // Gravenstein Highway North Sebastopol, California 95472, 2012. – 272 p.
4. Jana, A. Kinect for Windows SDK Programming Guide [Text] / A. Jana; Published by Packt Publishing Ltd. // Livery Place, 35 Livery Street, Birmingham B3 2PB, UK, 2012. – 392 p.
5. Catuhe, D. Programming with the Kinect for Windows [Text] / D. Catuhe // Microsoft Press, Redmond, Washington, 2012. – 224 с.
6. Кононович Л. М. Стерефоническое восприятие звука [Текст] / Л. М. Кононович, Ю. А. Ковалгин. – М.: Радио и связь, 1981. – 184 с.
7. Алдошина И. А. Электроакустика и звуковое вещание: Учебное пособие для вузов [Текст] / И. А. Алдошина, Э. И. Вологдин, А. П. Ефимов и др.; под ред. Ю. А. Ковалгина. – М.: Горячая линия-Телеком, Радио и связь, 2007. – 872 с.
8. Алдошина, И. А. Высококачественные акустические системы и излучатели [Текст] / И. А. Алдошина, А. Г. Войшвило. – М.: Радио и связь, 1985. – 168 с.
9. Бесекерский, В. А. Теория систем автоматического управления [Текст] / В. А. Бесекерский, Е. П. Попов; изд. 4-5, перераб. и доп. – СПб, Изд-во «Профессия», 2003. – 752 с.
10. Алябьев, С. И. Радиовещание и электроакустика [Текст] / С. И. Алябьев, А. В. Выходец, Р. Гермер и др. – М.: Радио и связь, 1998. – 783 с.
11. Виноградова, Э. Л. Конструирование громкоговорителей со сглаженными частотными характеристиками [Текст] / Э. Л. Виноградова. – М.: Энергия, 1978. – 48 с.
12. Шкритек, П. Справочное руководство по звуковой схемотехнике [Текст] / П. Шкритек; пер. с нем. – М.: Мир, 1991. – 446 с.
13. Волович, Г. И. Схемотехника аналоговых и аналого-цифровых электронных устройств [Текст] / Г. И. Волович. – М.: Издательский дом «Додэка-XXI», 2005. – 528 с.
14. Owsinski, B. The mixing engineer's handbook, Second Edition [Text] / B. Owsinski. – Thomson Course Technology, 2006. – 290 с.