Sociedade de Engenharia de Áudio

Artigo de Congresso

Apresentado no 12º Congresso de Engenharia de Áudio
18ª Convenção Nacional da AES Brasil
13 a 15 de Maio de 2014, São Paulo, SP

Este artigo foi reproduzido do original final entregue pelo autor, sem edições, correções ou considerações feitas pelo comitê técnico. A AES Brasil não se responsabiliza pelo conteúdo. Outros artigos podem ser adquiridos através da Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA, www.aes.org. Informações sobre a seção Brasileira podem ser obtidas em www.aesbrasil.org. Todos os direitos são reservados. Não é permitida a reprodução total ou parcial deste artigo sem autorização expressa da AES Brasil.

Sistema de Realidade Aumentada Áudio 3D para Dispositivos iOS

Sérgio I. Lopes,[1,2] José M. N. Vieira,[1,2] Guilherme Campos[1,2] e Paulo Dias[1,2]

[1] Universidade de Aveiro, Departamento de Eletrónica, Telecomunicações e Informática
3810-193 Aveiro, Portugal

[2] Instituto de Engenharia Electrónica e Telemática de Aveiro, 3810-193 Aveiro, Portugal

sil@ua.pt, jnvieira@ua.pt, guilherme.campos@ua.pt, paulo.dias@ua.pt

RESUMO

Neste trabalho é apresentado um sistema de realidade aumentada áudio 3D (binaural) para dispositivos móveis iOS. O sistema tira partido dos recursos de processamento e entrada/saída existentes num dispositivo iOS para obter a posição do utilizador recorrendo a um sistema de localização acústica e da unidade de medição inercial (IMU) existente no dispositivo para obter a sua orientação. Pretende-se que o utilizador tenha uma experiência de áudio 3D (i.e., áudio binaural com acústica de sala virtual) tendo em conta a sua posição e orientação relativa a fontes de áudio virtuais colocadas em pontos de interesse numa sala real. O principal objetivo é criar a ilusão de que um objeto funciona como uma fonte sonora instalada num ambiente acústico virtual que tirando partido da posição/orientação do utilizador para processar a fonte sonora gere uma saída binaural em tempo-real, que pode ser ouvida pelo indivíduo através de auscultadores. O sistema proposto representa a evolução do trabalho descrito em [1] sendo a localização do utilizador baseada no trabalho descrito em [2].

0 MOTIVAÇÃO

A principal motivação para este artigo é dar seguimento ao trabalho desenvolvido previamente em [1], cujo principal objetivo era o desenvolvimento de um sistema de realidade aumentada áudio 3D para museus. O sistema era capaz de localizar um utilizador e a sua orientação da cabeça numa sala real e virtualizar diversas fontes sonoras no espaço. Tendo em conta a posição/orientação relativa do utilizador a cada uma das fontes virtuais o sistema gerava em tempo-real uma versão binaural com o objetivo de criar a ilusão que as obras em exposição no museu (e.g., pinturas, esculturas ou outras obras de arte) emitiam som. Na versão apresentada em [1] o protótipo utilizava um sistema de

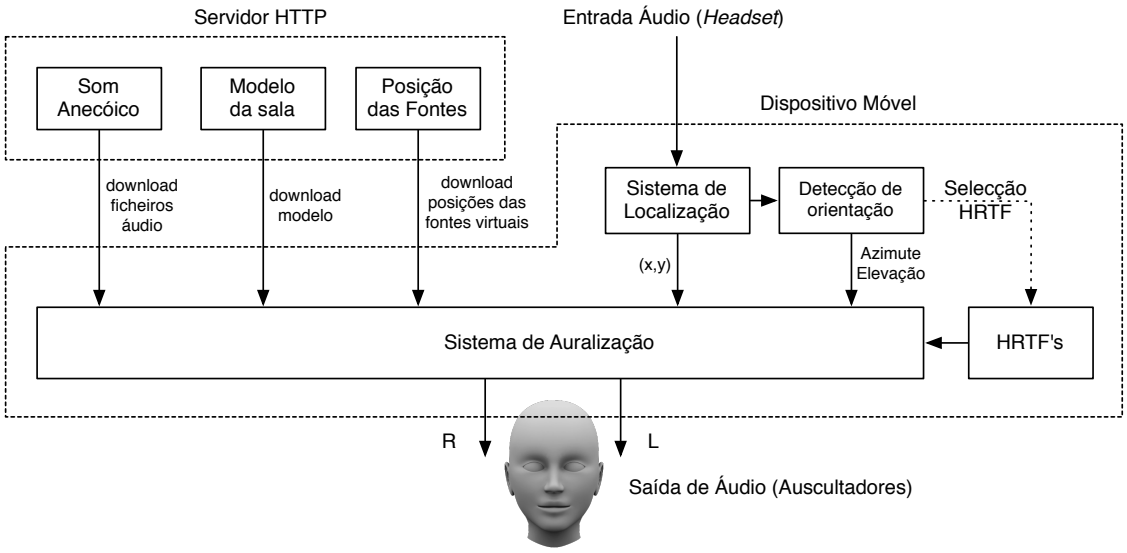

Figura 1: Diagrama de blocos do sistema proposto.

localização ultrasónico de banda larga [3] desenvolvido tinha por base um computador portátil o que inviabilizava a utilização do sistema em dispositivos móveis. Neste momento o sistema encontra-se a ser portado para iOS, faltando apenas o bloco de auralização construído a partir da biblioteca AAVE previamente descrita em [4].

1 ARQUITECTURA GERAL

A arquitetura geral do sistema está descrita na Figura 1, na qual se pode observar os seus módulos principais: módulo de localização e de orientação do utilizador e módulo de auralização.

O módulo de localização permite estimar a posição do utilizador na sala através da utilização de uma infra-estrutura de faróis acústicos sincronizados que transmitem sinais áudio não-invasivos de forma periódica. A utilização de uma infra-estrutura síncrona permite utilizar medidas de Diferença de Tempo de Chegada (TDoA) para estimar a posição do utilizador, libertando o dispositivo móvel de estar sincronizado com a infra-estrutura. Para além disso, com um funcionamento em *downlink* (i.e. os faróis funcionam como transmissores e os dispositivos móveis como receptores, o número de estações móveis é ilimitado) [5].

O bloco de auralização processa assim o som anecóico de uma fonte virtual para gerar um som binaural (3D) tendo em conta as entradas descritas na Figura 1. O seu funcionamento é baseado na convolução dos sinais monofónicos anecóicos emitidos por cada fonte sonora virtual com pares de filtros HRTF (*Head Related Transfer Functions*) descritos em [6] e [7], que são selecionados de acordo com a azimute e a elevação da cabeça do utilizador. A intensidade das fontes de áudio virtuais depende tanto da orientação da cabeça do utilizador (relação com a HRTF) como da distância do utilizador à fonte de som virtual, tendo sido aplicada a lei de atenuação do inverso do quadrado da distância a cada fonte sonora [8]. A orientação do indivíduo é feita tendo por base a unidade IMU disponibilizada pelo dispositivo sendo que para uma utilização realista o dispositivo iOS deverá estar fixo relativamente ao utilizador. Caso se pretenda detectar a orientação da cabeça do utilizador o dispositivo deverá ser acoplado aos auscultadores de forma a replicar o movimento da cabeça do utilizador. Em [9] os autores referem ser possível obter uma precisão menor que 15° na orientação de um dispositivo iOS, quando o utilizador se encontra a caminhar em ambientes interiores, podendo a performance melhorar para situações em que se encontra em repouso.

1.1 Localização do Utilizador

O sistema de localização funciona num modo de operação distribuída, ou seja, cada dispositivo iOS é capaz de estimar a sua própria posição usando uma topologia análoga à do GPS, em que os faróis operam como emissores e os dispositivos móveis como receptores. Além disso, o dispositivo móvel não precisa de estar sincronizado com a infra-estrutura de faróis uma vez que utiliza medidas de TDoA para obter uma estimativa de posição. O dispositivo iOS não precisa de nenhum hardware externo no processo de localização, apenas um vulgar microfone de eletreto ligado à entrada de microfone do *headset* que deve estar acoplado aos auscultadores permitindo deste modo a sua colocação mais próxima do ponto central ideal da cabeça do utilizador permitindo uma redução considerável das situações em que emissor e receptor não estão em linha-de-vista (NLoS). Para efetuar as medições TDoA são utilizados *chirps* pulsados com conteúdo de frequência no limite da gama do audível, na banda dos 18 kHz aos 22 kHz . Para minimizar o impacto dos pulsos transmitidos nos humanos a envolvente dos pulsos foi modu-

lada tendo em conta a redução da audibilidade dos pulsos e a maximização da transferência de energia (ver detalhes de implementação em [2]), de forma a poder tirar o máximo partido da resposta em frequência da entrada de microfone do *headset* existente nos dispositivos iOS, ver Figura 2.

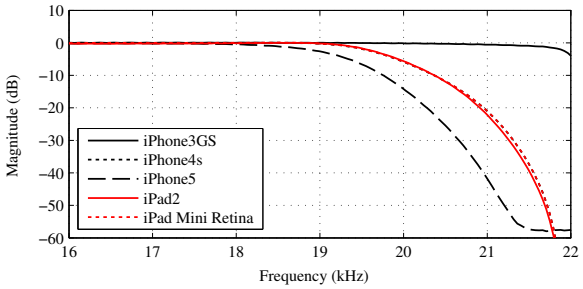

Figura 2: Resposta em frequência da entrada de microfone (*headset*) em dispositivos iOS.

1.2 Auralização

Este bloco tem como função gerir a posição das fontes sonoras num ambiente áudio virtual relativamente à posição e orientação do ouvinte. O processamento é feito utilizando o método das imagens virtuais para o cálculo das reflexões do som nas superfícies do modelo geométrico de uma sala virtual. Em [4] é descrito em detalhe o bloco de auralização tendo resultado no desenvolvimento da biblioteca AAVE. A componente geométrica do módulo de auralização AAVE está assim dependente de dados externos, tal como apresentado na Figura 1 que do ponto de vista da aplicação estão alojados num servidor remoto. Deste modo é possível configurar o sistema para diferentes modelos de sala e diferentes fontes sonoras e respectivas posições.

2 Implementação em iOS

Todas as entradas/saídas são feitas utilizando os recursos disponíveis no dispositivo iOS. A entrada de microfone do *headset* é utilizada como entrada para o sistema de localização e a saída *stereo* do mesmo utilizada para ligar os auscultadores do utilizador. Tendo em conta os requisitos da aplicação optou-se por fixar o microfone nos auscultadores. Desta forma, foi possível colocar um microfone externo no topo da cabeça do utilizador e tirar partido da unidade de medição inercial (IMU) para obtenção da informação de orientação.

Nas Figuras 3 e 4 estão representados os diagramas de blocos simplificados do *software* desenvolvido, onde se podem observar as duas rotinas de serviço à interrupção (RSI) utilizadas. A RSI relativa à entrada de áudio, ver diagrama da Figura 3, está completamente alocada ao processamento crítico necessário ao funcionamento do sistema de localização, desde a leitura do *buffer* de entrada e filtragem ótima seguindo-se a estimação de TDoA e a consequente localização hiperbólica do utilizador.

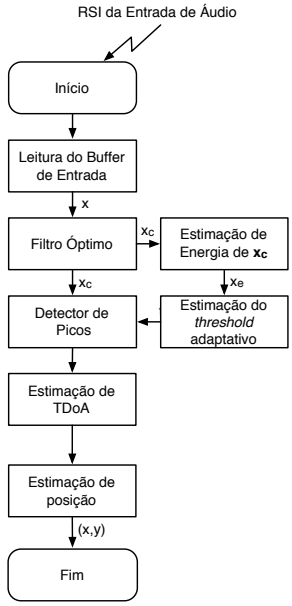

Figura 3: Rotina de serviço à interrupção da entrada de áudio (localização) no dispositivo iOS.

Por outro lado na RSI relativa à saída de áudio, ver diagrama da Figura 4, é efetuado o processamento crítico necessário ao funcionamento do sistema de auralização. Nesta rotina é obtida a informação de orientação do dispositivo e tendo em conta a posição atual do utilizador, o modelo acústico da sala e a informação relativa às fontes sonoras é efetuada a auralização do sinal anecóico a partir da correta seleção de um par de HRTF's de uma base de dados local.

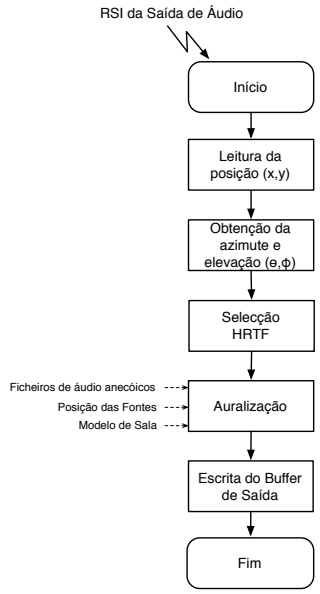

Figura 4: Rotina de serviço à interrupção da saída de áudio (auralização) no dispositivo iOS.

a) b) c)

Figura 5: **a)** Protótipo do dispositivo móvel com auscultadores equipados com microfone e IMU. **b)** *Interface* principal da aplicação. **c)** *Interface* de *debug* para desenvolvimento.

A auralização é posteriormente efetuada tendo em conta informação vinda do servidor e guardada localmente no dispositivo iOS no processo de inicialização do sistema, tal como o modelo da sala e a localização das fontes na sala.

Todos os algoritmos do sistema de localização (detecção de pulsos, estimação TDoA e posicionamento hiperbólico) foram programados utilizando Objective-C++, tendo por base a *Accelerate framework* existente para iOS que permite executar processos de filtragem no domínio do tempo (convolução) e operações baseadas na FFT de uma forma eficiente. Esta *framework* utiliza o processador GPU existente no dispositivo para aumentar o desempenho computacional tirando partido da sua arquitetura desenvolvida para operações orientadas a matrizes e vectores. A obtenção da orientação do dispositivo é feita através da *Core Location framework* que permite obter a orientação do dispositivo (a partir da informação sensorial dada pelo acelerómetro, giróscópio e magnetómetro) existentes no dispositivo. Na Figura 6 é apresentado o resultado da medição do erro acumulado ao longo do tempo na obtenção da orientação quando é utilizado um iPhone 4s em constante rotação em torno do eixo z em ambiente interior. Analisando a figura, podemos observar um *drift* elevado (aproximadamente 20° por minuto) quando é apenas utilizada a informação do sensor IMU. A inclusão da informação do magnetómetro no algoritmo de fusão sensorial revela-se eficaz reduzindo este *drift* para um erro acumulado abaixo dos 2°, tornando possível a utilização do mesmo na aplicação em questão.

3 PROTÓTIPO

Na Figura 5 é apresentado o protótipo do dispositivo móvel que inclui os auscultadores com o microfone utilizado no sistema de localização. Na imagem (b) podemos observar a *interface* principal da aplicação desenvolvida para iOS onde se podem visualizar a posição atual do utilizador (círculo amarelo) a posição dos faróis acústicos na sala (círculos azuis) e a posição das fontes sonoras virtuais (círculos vermelhos) com as respectivas zonas de ação delimitadas. Na imagem (c) podemos observar a *interface* de *debug* utilizada para o desenvolvimento que permite monitorizar em tempo-real o sinal de entrada depois de passar pelo filtro ótimo bem como outras características relacionadas com o sinal que está a ser utilizado para além de informação relacionada com o método de partilha do meio utilizado.

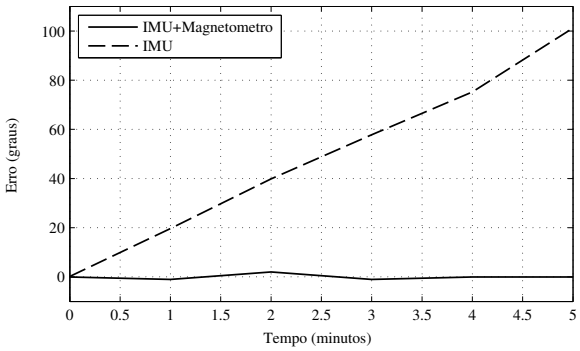

Figura 6: Erro acumulado na medição de orientação quando o dispositivo (iPhone 4s) está em rotação sob o eixo z.

4 CONCLUSÕES E TRABALHO FUTURO

Na Figura 7 é apresentado um exemplo de um percurso com dados reais da localização do utilizador alcançados com um iPhone 4s tendo sido obtido um erro absoluto inferior a 16 cm em 95% dos casos para

um período de atualização de posição de 350 ms. A medição da orientação do utilizador apresenta um erro acumulado menor que 2° quando o dispositivo está em constante rotação em torno do eixo z em ambiente interior. Nos testes efetuados com o iPhone 4s, apenas com o sistema de localização/orientação obtivemos uma taxa de ocupação do processador de cerca de 40% o que irá limitar o número total de fontes sonoras ativas (incluindo as fontes-imagem) a utilizar no processo de auralização com a biblioteca AAVE na sua versão iOS. Nesta fase, quer o modelo acústico da sala, quer as fontes virtuais e suas posições já estão disponíveis na aplicação estando a ser portada a biblioteca AAVE para iOS.

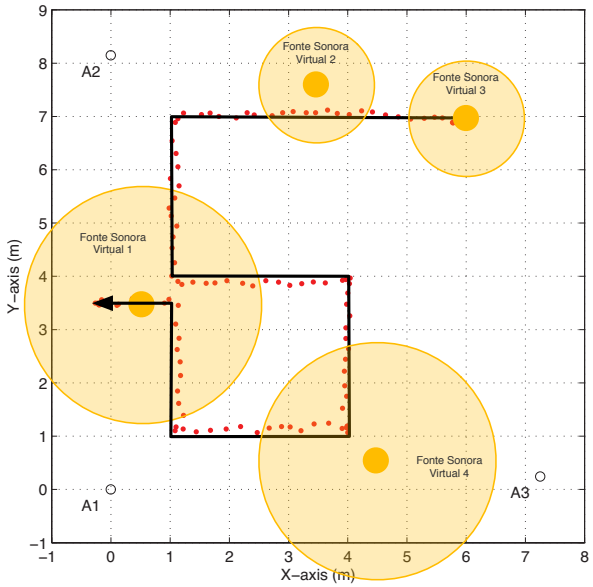

Figura 7: Exemplo com dados reais de um percurso de um utilizador já com fontes sonoras virtuais adicionadas. A1, A2 e A3 representam faróis acústicos necessários ao sistema de localização.

AGRADECIMENTOS

Este trabalho foi financiado pelo Estado Português através da FCT - Fundação para a Ciência e a Tecnologia no âmbito do projeto AcousticAVE: Modelos e Aplicações de Auralização em Ambientes de Realidade Virtual (PTDC/EEA-ELC/112137/2009). Este trabalho foi também financiado pelo FEDER - Fundo Europeu de Desenvolvimento Regional através do COMPETE - Programa Operacional Factores de Competitividade e pelo Estado Português através da FCT - Fundação para a Ciência e a Tecnologia no âmbito do projeto FCOMP-01-0124-FEDER-022682 (FCT PEst-C/EEI/UI0127/2011).

REFERÊNCIAS BIBLIOGRÁFICAS

[1] S.I. Lopes, A. Oliveira, J.M.N. Vieira, G. Campos, P. Dias, and R. Costa, "Real-time audio augmented reality system for pervasive applications," in *Proceedings of the 11th Audio Engineering Society Brazil Conference*, R.R.A. Faria, Ed., Sao Paulo, Brazil, 7-9 May 2013, AES Brazil, ISSN 2177-529X.

[2] S.I. Lopes, J.M.N. Vieira, J. Reis, D.F. Albuquerque, , and N.B. Carvalho, "Accurate smartphone indoor positioning using non-invasive audio," in *Proceedings of the 2013 International Conference on Indoor Positioning and Indoor Navigation - IPIN 2013*, Montbéliard - Belfort, France, 28-31st October 2013.

[3] S.I. Lopes, Vieira, and D. Albuquerque, "High Accuracy 3D Indoor Positioning using Broadband Ultrasonic Signals," in *Trust, Security and Privacy in Computing and Communications (TrustCom), 2012 IEEE 11th International Conference on*, 2012, pp. 2008–2014.

[4] André Oliveira, Guilherme Campos, Paulo Dias, Damiam Murphy, José Vieira, Catarina Mendonça, and Jorge Santos, "Real-time dynamic image-source implementation for auralisation," in *Proc. of the 16th Int. Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, September 2-6 2013.

[5] Seyed A. (Reza) Zekavat and R. Michael Buehrer, *Handbook of Position Location - Theory, Practice, and Advances*, John Wiley & Sons, Inc., 2011.

[6] Corey Cheng and Gregory Wakefield, "Introduction to Head-Related Transfer functions (HRTFs): Representation of HRTFs in Time, Frequency and Space," *J. Audio Eng. Soc. (AES)*, vol. 49, no. 4, 2001.

[7] Bill Gardner and Keith Martin, "HRTF Measurements of a KEMAR Dummy-Head Microphone," 1994, http://sound.media.mit.edu/resources/KEMAR.html.

[8] David M. Howard and James A. S. Angus, *Acoustics and Psychoacoustics*, Focal Press, Elsevier, fourth edition, 2009.

[9] JeffreyR. Blum, DanielG. Greencorn, and JeremyR. Cooperstock, "Smartphone sensor reliability for augmented reality applications," in *Mobile and Ubiquitous Systems: Computing, Networking, and Services*, Kan Zheng, Mo Li, and Hongbo Jiang, Eds., vol. 120 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, pp. 127–138. Springer Berlin Heidelberg, 2013.