

Tecnologia de Captura de Movimento Facial Aplicada ao Estudo de Padrões Articulatorios da Fala

Mateus A. Chinelatto*, Paula D. P. Costa

Resumo

A tecnologia de captura de movimento, ou MoCap, encontra grande aplicação na realização de estudos biomecânicos em esportes e na indústria do entretenimento para a criação de personagens com movimentação mais realista em filmes e jogos digitais. O presente trabalho utilizou equipamento especializado na captura de movimentos faciais para o estudo e identificação de padrões articulatorios da fala, visando sua modelagem para sistemas automáticos de animação facial 3D. Utilizando um equipamento e tecnologias mais sofisticados, o presente trabalho revisita estudo anteriormente realizado no DCA/FEEC, que identificou os principais visemas dependentes de contexto fonético para o português do Brasil.

Palavras-chave:

captura de movimento, animação facial, clusterização.

Introdução

Para a geração de animações faciais videorealistas é muito importante modelar a fala de maneira adequada, sendo assim, é essencial a identificação de seus padrões articulatorios. Isso é feito identificando os visemas de determinada língua. Visemas são os equivalentes visuais dos sons produzidos pela fala da língua em questão (fonemas).

Em 2005 De Martino identificou os principais visemas dependentes do contexto fonético para o português do Brasil, propiciando uma melhora na qualidade dos sistemas de animação facial voltados para a língua portuguesa [1]. Neste projeto o objetivo foi rever esse trabalho de identificação dos visemas explorando a técnica de MoCap, por meio de um equipamento mais sofisticado, capaz de rastrear a posição tridimensional de pontos marcados no rosto de um ator.

Resultados e Discussão

Dentre os resultados alcançados por este trabalho está a criação de uma base de dados de capturas de movimentos faciais constituída por 102 capturas. Destas, 81 são constituídas por palavras paroxítonas de duas sílabas, formadas por um som vocálico seguido de uma vogal e outras 21 por apenas uma sílaba formada com duas vogais [1].

Raramente uma captura é perfeita, normalmente existem falhas na detecção de alguns pontos, conflitos entre os dados de cada câmera do equipamento, por exemplo pontos diferentes sendo rastreados como sendo o mesmo ponto e rastreamento de pontos indesejados como as narinas ou as pupilas (Figura 1). Isso exigiu um processamento individual de cada captura para corrigir essas falhas e constituir uma base de resultados confiáveis.

O equipamento de MoCap não realiza a gravação de áudio e por isso foi utilizado um gravador externo. Juntamente com a base das capturas de movimento, foi construída uma base de dados com os áudios referentes a elas e suas transcrições fonéticas. Essas transcrições são arquivos que descrevem os instantes em que cada fonema começa a ser emitido e sua duração. Isso tornou possível associar cada som a uma sequência de frames (quadros) capturados.

Além das bases de dados citadas e de todo o processamento, outro resultado do projeto foi a identificação do visema que representava cada som gravado. Isso foi feito identificando o quadro correspondente ao maior deslocamento dos pontos rastreados em relação a sua posição de repouso, que seria quando o ator está em silêncio.

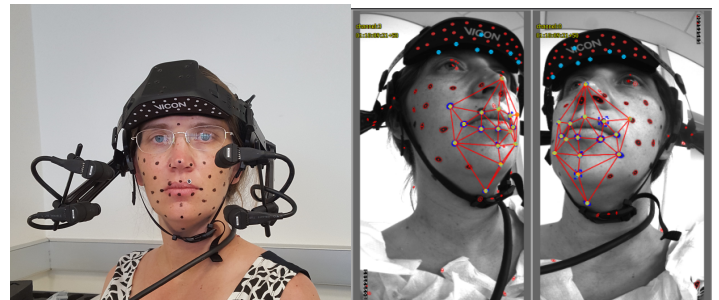


Figura 1. Equipamento de captura de movimento com pontos marcados na face e exemplo de processamento mostrando as imagens das câmeras inferiores.

Conclusão

Os resultados obtidos neste trabalho visam modelar de forma adequada a fala do português do Brasil, com a finalidade de melhorar ainda mais a qualidade e videorealismo dos sistemas de animação facial. Baseando a metodologia e a análise em métodos mais precisos é de se esperar que isso seja alcançado. Porém, é necessária ainda uma análise mais profunda que identifique similaridades entre os visemas encontrados, agrupando-os por meio de algoritmo de clusterização e comparando os grupos encontrados com os obtidos por De Martino [1].

O próximo passo seria a validação dos resultados por meio da proposta de um sistema de animação facial 3D e por fim teste de inteligibilidade da fala.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio financeiro a este projeto.

[1] DE MARTINO, José Mario et al. *Animação facial sincronizada com a fala: visemas dependentes do contexto fonético para o português do Brasil*. 2005.