

**Universidade do Porto**  
**Faculdade de Engenharia do Porto**  
Mestrado em Tecnologias Multimédia

**CARTOON** *broadcast system*

Luís Miguel Barbosa da Costa Leite  
FEUP – 2005.2006

<a href="#"><u>1. Introdução</u></a> .....	3
<a href="#"><u>2. Estado da Arte</u></a> .....	5
<a href="#"><u>2.1. Reconhecimento da fala</u></a> .....	5
<a href="#"><u>2.2. Motion capture</u></a> .....	7
<a href="#"><u>2.2.1. Tecnologia</u></a> .....	7
<a href="#"><u>2.2.2. Sistemas Mecânicos</u></a> .....	8
<a href="#"><u>2.2.3. Sistemas Magnéticos</u></a> .....	9
<a href="#"><u>2.2.4. Sistemas Ópticos</u></a> .....	10
<a href="#"><u>2.2.5. Sistemas Acústicos</u></a> .....	13
<a href="#"><u>2.3. Face capture</u></a> .....	14
<a href="#"><u>3. Conclusão</u></a> .....	15

## 1. Introdução

O *Cartoon Broadcast System* é uma proposta para criar uma plataforma técnica que suporte animações (desenhos animados) interactivas em tempo real para emissão ou projecção, como se tratasse de um teatro de marionetas digital, mas num formato mais televisivo.

Este estúdio de animação deverá oferecer ao realizador a possibilidade de escolher câmaras, personagens, animações pré-definidas e inseri-las num *timeline*, por outro lado o sistema permitirá o sincronismo labial das personagens com os actores de voz, a manipulação das personagens e o controle das suas expressões de forma a ser possível misturar e criar um espectáculo de animação ao vivo pronto a ser emitido ou projectado. A interface deste sistema contemplará a parte da realização, com múltiplas câmaras, personagens e cenários. A manipulação das personagens será constituída com interfaces de entrada de dados (*inputs*), tais como: teclados, joysticks, ratos, sistemas midi, mesas digitalizadoras, sistemas de captura de movimento (*face capture* e *motion capture*) e a própria voz.

A presente proposta tem como público-alvo os canais de televisão, as produtoras, performers e/ou escolas.

Para desenvolver este sistema é necessário um estudo aprofundado sobre sistemas semelhantes, ou seja, tecnologias que possam ser integradas, tais como técnicas de animação, vídeo e áudio que possam ajudar a melhorá-lo. É também necessário um estudo sobre uma interface que sirva todos os propósitos.

Da investigação efectuada, com o objectivo de procurar sistemas semelhantes, surgiu apenas um com as mesmas características e os mesmos propósitos: um sistema desenvolvido pela *Tv-animation* e utilizado no famoso programa *Nelly Nuts* transmitido por várias cadeias de televisão (BBC, Fox Kids). Curiosamente, Portugal foi o primeiro país a adquirir este sistema ainda em fase de testes. Foi comprado pela produtora Miragem que adaptou o conceito *Nelly Nuts* para *Rita Catita*, o qual foi transmitido em directo pela TVI. Tive a oportunidade de acompanhar de um modo bastante implicado a implementação do sistema nos estúdios da TVI e de trabalhar directamente com o realizador responsável, Nuno Carvalho, ao qual dirigi já o convite para participar no desenvolvimento deste novo sistema.

A plataforma da *Tv-animation* envolvia dois computadores de *broadcast*, um para a emissão das animações comandadas pela realização e controlo do sincronismo labial,

um outro para a emissão dos jogos interactivos; as duas placas de emissão estavam ligadas a uma mesa de mistura vídeo que também recebia o sinal de sincronismo (timecode) gerado pelo primeiro computador de emissão; existia ainda um terceiro computador que servia como interface de realização, onde se podia controlar, alguns presets de animação, montar as cenas e picar as câmaras em tempo real. Estes computadores estavam ligados a um hub por rede ethernet permitindo a comunicação entre todos os computadores.

A plataforma da *Tv-animation*, apesar de se ter transformado num sucesso em vários países, denotava uma grande insegurança e alguma apatia em termos de evolução e resolução de problemas.

Na presente proposta gostaria de assinalar algumas diferenças de funcionamento em relação à plataforma da *Nelly Nuts*: a introdução de manipuladores das marionetas, detectores de expressões das personagens, inserção de áudio referencial de determinadas acções, tornar o sistema modular podendo acoplar novos módulos de manipulação de personagens automaticamente, adoptar uma linguagem de comunicação midi para ligação de múltiplas peças de hardware para controlo de várias funções, introdução de cenários croma keyer para poder misturar com imagens reais de forma a permitir intervenções com o mundo real e em directo.

Existem no entanto outros sistemas que também servem para *broadcast*, mas que normalmente são utilizados com a função de pivots virtuais ou cenografia virtual.

É o caso da Kaydara FilmBox, recentemente comprada pela Alias, que produziu o programa de apresentação de desenhos animados com o Johnny Bravo Cartoon para a Cartoon networks. Neste caso foi utilizado uma versão beta do FilmBox on-air, e o Filmbox Voice Reality para o sincronismo labial.

Existem muitos outros sistemas de *broadcast* de personagens em tempo real, como o Alias Motion Builder para personagens 3D, ou o ORAD system, mas normalmente servem apenas para colocar uma personagem animada em tempo real sobre um cenário real ou virtual e não são sistemas completos como a plataforma da *Tv-animation*, ou como a proposta apresentada.

Várias tecnologias podem ser utilizadas para o desenvolvimento desta plataforma, como o *motion capture*, o *face capture*, o reconhecimento da fala, ou outros sistemas de interacção.

Desenvolvi estudos sobre algumas das tecnologias mais pertinentes para saber o estado da arte e também para poder iniciar algumas opções para este sistema.

## 2. Estado da Arte

### 2.1. Reconhecimento da fala

Algumas das principais áreas de aplicação comercial para os sistemas de reconhecimento automático de fala são: ditado, interfaces para computadores pessoais, serviços de atendimento de chamada automáticos, aplicações industriais especiais. A principal razão para o sucesso comercial tem sido o aumento na produtividade proporcionado por estes sistemas, os quais auxiliam ou substituem operadores humanos. As tecnologias mais utilizadas são: HMMS, MLPS, MFCCS.

Nos sistemas que constituem o estado da arte na área de reconhecimento de fala predominam os modelos estatísticos, notoriamente aqueles baseados em Modelos Ocultos de Markov (*Hidden Markov Models*, HMM). Os HMM's são estruturas poderosas pois são capazes de modelar ao mesmo tempo as variabilidades acústicas e temporais do sinal de voz.

Nos sistemas de reconhecimento da fala contínua e vocabulário extenso, as palavras são geralmente modeladas a partir da concatenação de subunidades fonéticas, pois o número destas é bem menor do que o de palavras, e numa locução geralmente existem vários exemplos de subunidades fonéticas.

O reconhecimento da fala contínua difere do de palavras isoladas, pois neste o locutor não precisa de fazer pausas entre as palavras. Deste modo, a determinação das fronteiras entre as palavras e do número destas na locução deve ser feita pelo sistema de reconhecimento. Para isto são utilizados os algoritmos de busca, que podem ter ainda modelos de duração e de linguagem incorporados.

A última década tem testemunhado um progresso significativo na tecnologia de reconhecimento de fala. As taxas de erro de palavra caem de um factor de 2 a cada dois anos. Foi feito um progresso substancial na tecnologia básica, o que levou a vencer a barreira da independência do locutor, fala contínua e vocabulários extensos. Existem vários factores que contribuíram para este rápido progresso.

A chegada da era do HMM. O HMM é poderoso no sentido de que, com a disponibilidade de dados da aprendizagem da voz, os parâmetros do modelo podem ser regulados automaticamente para fornecer um melhor desempenho.

Foi feito um grande esforço no sentido de desenvolver grandes bases de dados de fala para o desenvolvimento, treino e avaliação de sistemas de reconhecimento da fala, de forma ao estabelecimento de normas de avaliação de desempenho.

A recente disponibilidade de grandes bases de dados no domínio público, associada à especificação de padrões de avaliação, resultou numa documentação uniforme de resultados de testes, contribuindo para uma maior fiabilidade na monitorização dos progressos alcançados.

Os avanços na tecnologia dos computadores influenciaram indirectamente o progresso nesta área. A disponibilidade de computadores rápidos com grandes capacidades de memória permitiu aos investigadores realizar várias experiências em larga escala e em um curto espaço de tempo. Isto significa que o intervalo de tempo entre uma ideia e a sua implementação e avaliação foi bastante reduzido. De facto, sistemas de reconhecimento de fala com desempenho razoável podem ser executados em microcomputadores comuns em tempo real, sem hardware adicional, um facto inimaginável à alguns anos atrás.

As taxas de erro variam consoante o ambiente acústico, as características do orador, o número de palavras e as tecnologias utilizadas.

Talvez a tarefa mais popular, com baixa perplexidade é o reconhecimento de dígitos conectados. Para o inglês, o reconhecimento independente do locutor de sequências de dígitos pronunciados de forma contínua e restringido à largura de banda telefónica pode alcançar uma taxa de erro de 0,3% quando o comprimento da sequência é conhecido.

Uma das tarefas de média perplexidade mais conhecidas é a de 1000 palavras, chamada de *Resource Management*. O melhor desempenho independente do locutor nesta tarefa é de menos de 4%, utilizando um modelo de linguagem de pares de palavras que limita as palavras possíveis que seguem uma dada palavra. Recentemente, os investigadores começaram a estudar a questão do reconhecimento de fala espontânea.

Tarefas com alta perplexidade, com vocabulários de milhares de palavras, são destinadas principalmente para aplicações de ditado. Depois de trabalhar em sistemas de palavras isoladas, dependentes do locutor, a comunidade científica voltou-se para o reconhecimento de fala contínua para grandes vocabulários (20.000 palavras ou mais), alta perplexidade, independente do locutor. Já em 1997 o melhor sistema alcançava uma taxa de erro de 9,9% em testes realizados regularmente nos EUA através do Departamento de Defesa.

Actualmente, os sistemas que representam o estado da arte em reconhecimento de fala contínua baseiam-se em Modelos Ocultos de Markov, capaz de modelar tanto as variabilidades acústicas como temporais do sinal de fala. Porém, para viabilizar a modelação matemática de um HMM, são realizadas inúmeras suposições simplificadoras que limitam o seu potencial efectivo. Redes neuronais artificiais – ANN (“Artificial Neural Networks”) que não necessitam fazer o uso de muitas destas suposições, podem aprender e generalizar superfícies complexas de decisão, tolerar ruídos e suportar paralelismo. Todas estas vantagens tornam as ANNs extremamente poderosas para modelar as variabilidades acústicas da fala. Entretanto, ao contrário dos HMMs, as ANNs não se têm mostrado eficientes para a modelação das variabilidades temporais. Com o objectivo de unir numa única estrutura o que há de melhor nas tecnologias de redes neuronais artificiais e de Modelos Ocultos de Markov, têm sido estudados e avaliados nos últimos anos, modelos híbridos ANN-HMM.

Um novo horizonte de investigação se deslumbra no reconhecimento da fala com a fusão da informação recolhida da imagem (aplicando a DCT2) e do sinal acústico.

Depois de vários anos de pesquisa, a tecnologia de reconhecimento de fala está a passar o limiar da praticabilidade. Na última década assistiu-se a um grande progresso na tecnologia de reconhecimento de fala, no sentido da disponibilidade de algoritmos e sistemas de alto desempenho. Em muitos casos, a transição de protótipos de laboratório para sistemas comerciais já se iniciou.

## ***2.2. Motion capture***

A técnica de *Motion capture* consiste em capturar a posição e orientação de certas partes de um objecto real (no caso de um ser humano, em geral as articulações), utilizando processos que podem ser ópticos, mecânicos, electromagnéticos ou acústicos.

### **2.2.1. Tecnologia**

Podemos dividir os sistemas de captura em quatro categorias: mecânicos, ópticos, magnéticos e acústicos. Actualmente, os sistemas mais utilizados são baseados na tecnologia óptica e magnética.

Estas tecnologias são normalmente acompanhadas por várias aplicações de controlo, de sincronismo, de análise, de correcção de erros e optimização. Um exemplo de uma destas aplicações é o CLIMA (Clinical Motion Analysis). O CLIMA captura o movimento, faz uma análise biomecânica, e mostra todos os resultados quantificados,

como os dos ângulos das articulações e das velocidades, as posições, e as acelerações, centro de gravidade, permitindo a comparação com valores normais.

### **2.2.2. Sistemas Mecânicos**

Os sistemas mecânicos são compostos por potenciômetros que são posicionados nas articulações desejadas. Estes fornecem a informação sobre as coordenadas das posições e orientações com altas taxas de amostragem (tempo real).

Os sistemas mecânicos não recorrem a uma tecnologia muito sofisticada, porém possuem algumas vantagens que os tornam extremamente atraentes, especialmente para a indústria cinematográfica.

Sistemas mecânicos são equipamentos de medida absoluta, não sendo afectados por campos magnéticos ou reflexões indesejadas. Consequentemente, não necessitam de um processo longo de calibragem, o que torna a sua utilização mais fácil e produtiva.

Como exemplos de sistemas mecânicos de captura, podemos citar o *The Monkey – Digital Poseable Mannequin* e o *Animatton*. O primeiro pode ser considerado como um “*stop-motion*” assistido por computador; neste sistema, o “manequim” pode ser qualquer objecto desejado – desde humanóides até dinossauros e insectos. Um esqueleto articulado composto por potenciômetros que é fixado no modelo, fornecendo as posições e orientações. Uma desvantagem deste sistema está relacionada com o realismo, o qual depende exclusivamente da habilidade do animador em posicionar cada articulação do modelo correctamente, num processo que requer, além de técnica, um pouco de arte.

Os dados resultantes deste sistema traduzem-se em *keyframes* (fotogramas chave) da animação para posterior interpolação.

Uma nova tecnologia em sistemas mecânicos de captura foi desenvolvida pela Analagus, com o sistema *Animatton*. Este sistema, desenvolvido exclusivamente para a captura de movimentos humanos, é composto por articulações e potenciômetros que são fixados directamente no corpo do actor, formando uma verdadeira “armadura”. A captura é feita em tempo real e o sistema requer pouca ou nenhuma calibragem. Além disso, permite a captura de movimentos de múltiplos actores numa mesma sessão. A grande vantagem deste sistema é o preço, o que o torna acessível para a maioria das produtoras que utilizam o movimento capturado.

A resolução espacial deste sistema independente do alcance é de 0.5 a 1 graus com uma precisão espacial de 5 graus e precisão temporal superior a 120 fotogramas por segundo,

o espaço de trabalho é ilimitado. Animatton, Digital Monkey, Power Glove são alguns exemplos deste sistema capazes de funcionar com aplicações como os da Alias|Wavefront Maya, Discreet 3dStudio Max e Softimage XSI.

### **2.2.3. Sistemas Magnéticos**

Os sistemas magnéticos de captura caracterizam-se pela velocidade em tempo real de processamento dos dados capturados. Neste sistema emprega-se um conjunto de receptores que são posicionados nas articulações do actor. Estes receptores medem a posição e orientação das articulações no espaço, em relação a uma antena transmissora que emite um sinal.

A tecnologia dos sistemas magnéticos é menos sofisticada e os preços são mais acessíveis do que os sistemas ópticos. Existem sistemas magnéticos simples (Polhemus *InsideTrak*) e sofisticados (Ascension *Flock of Birds*). O preço deste tipo de sistema depende basicamente do número de receptores que conseguem ser processados simultaneamente.

Os sistemas magnéticos mais conhecidos são o *Ultratrak Pró*, da Polhemus e o *Flock of Birds*, da Ascension. Neste último, por exemplo, um conjunto até 30 receptores são conectados ao actor, fornecendo a orientação e posição das principais articulações.

Algumas vantagens dos sistemas magnéticos são o baixo custo computacional para o processamento dos dados, maior precisão (não existem problemas de oclusão de receptores) e baixo custo de equipamento. Com uma taxa de amostragem que ronda os 100 fotogramas por segundo, os sistemas magnéticos são ideias para a captura de movimentos mais simples.

A maior desvantagem deste tipo de equipamento está relacionada com os diversos cabos que ligam os receptores à antena. Tais cabos restringem o movimento do actor, não permitindo que movimentos complexos e rápidos possam ser representados com naturalidade. No entanto, algumas empresas já desenvolveram sistemas magnéticos sem cabos.

Uma outra desvantagem do processo magnético é a interferência causada por objectos de metal próximos ao local de captura. Campos magnéticos são extremamente sensíveis a objectos metálicos.

A resolução espacial deste sistema é de 0.005 a 8 mm e de 0.025 a 0.1 graus com uma precisão espacial de 0.8 a 5 mm e 0.1 a 3 graus e precisão temporal elevada, o espaço de trabalho é limitado ao raio de 1 a 2 metros. *Flock of Birds*, *Ultratrak* são alguns

exemplos deste sistema capazes de funcionar com aplicações como os da Alias|Wavefront Maya, Discreet 3dStudio Max e Softimage XSI.

#### **2.2.4. Sistemas Ópticos**

Neste tipo de sistema são utilizados reflectores (em geral, círculos ou esferas com material super-reflectivo), posicionados nas principais articulações (Figura 7). Câmaras especiais são posicionadas estrategicamente para fazer o *tracking* dos reflectores durante o movimento. Cada câmara gera coordenadas bidimensionais para cada reflector (obtidas via processo de segmentação). O conjunto de dados 2D capturados pelas câmaras independentes é então analisado por software que, através de um processo de triangulação, fornecerá as coordenadas tridimensionais dos reflectores.

Existem algumas variações entre diferentes sistemas ópticos de captura. No sistema *MultiTrax*, da Adaptive Optics Associates, os marcadores posicionados nas articulações são reflectores que, iluminados por flashes sincronizados por luz infra-vermelha, são capturados por uma ou mais câmaras. Neste tipo de equipamento óptico também podem ser utilizados marcadores de outros tipos (focos de laser e retro-prismas).

O sistema *Optotrak*, da Northern Digital, utiliza marcadores LED sincronizados, que são rastreados por três câmaras sensíveis ao espectro infra-vermelho. O software instalado no sistema utiliza os dados fornecidos pelas câmaras para gerar as coordenadas tridimensionais, orientações e momentos dos marcadores.

O sistema *ELITE*, da Bioengineering Tech., utiliza marcadores reflectivos passivos (hemisférios de plástico cobertos com material reflectivo) e câmaras electrónicas CCD com LED's infra-vermelhos que circulam a lente. O software instalado utiliza algoritmos de reconhecimento de formas (do tipo *Shape from shading*) para auxiliar o processo de *tracking* dos marcadores. Existem ainda muitos outros sistemas ópticos no mercado. Entre eles, podemos destacar o *HiRES ExpertVision 3D* (Motion Analysis Corp) e o *MacReflex* (Optikon Corp.).

Os equipamentos ópticos de captura são os mais caros do mercado. O preço elevado é o resultado da utilização da alta tecnologia, como câmaras de alta resolução.

Uma das vantagens da utilização destes sistemas é a alta taxa de amostragem, que permite a captura de movimentos rápidos como os utilizados em artes marciais, ou em alguns desportos (uma tacada de golfe). A taxa de amostragem depende basicamente da capacidade de definição e dos fotogramas por segundo das câmaras. As taxas de

amostragem chegam a atingir os 250 fotogramas por segundo com a utilização destas câmaras de alta resolução e velocidade.

Outra vantagem dos sistemas ópticos é a liberdade que oferecem ao actor durante a execução do movimento. Ao contrário dos sistemas magnéticos, onde o actor é coberto por receptores e cabos, os reflectores não oferecem resistência aos movimentos do actor. Além disso, não existe limite para o número de reflectores posicionados no actor, o que permite alcançar um maior número de detalhe na representação dos movimentos. O espaço de trabalho (*workspace*) do processo de captura para sistemas ópticos é muito superior ao dos sistemas magnéticos e depende, basicamente, do campo de visão (FOV) das câmaras utilizadas. A iluminação global ambiente deve ser calibrada de modo a não causar interferências no processo de *tracking* dos reflectores.

A grande desvantagem dos sistemas ópticos é a oclusão de um ou mais reflectores durante o processo de captura. Este tipo de problema é mais frequente durante a captura de movimentos de objectos pequenos (dedos das mãos) ou quando vários actores se cruzam. Neste caso, a recuperação da posição tridimensional do reflector pode ser impossível, o que compromete todo o processo.

O problema da oclusão pode ser minimizado com a utilização de um número maior de câmaras e de reflectores. Recorrer a mais câmaras acarreta um maior tempo de processamento durante o processo de *tracking*. Ao aumentar o número de reflectores, surge o problema de “*tracking confusion*”, ou seja, a dificuldade de identificar os reflectores que estão mais próximos. Este problema é directamente influenciado pela resolução das câmaras – câmaras com maior resolução conseguem identificar melhor reflectores próximos.

Existe ainda a necessidade de um processamento via software dos dados obtidos pelas câmaras para posterior utilização final, o que não permite muita interactividade, importante para a velocidade de produção e redução de custos operacionais. A utilização do software para o processamento dos dados obtidos introduz alguns erros de precisão no processo de obtenção de dados. Em alguns casos, é necessário um processo de filtragem dos dados finais para a eliminação dos erros de precisão e dos ruídos introduzidos pelo *tracking* das câmaras.

Uma empresa que tem dado provas com os seus equipamentos de captura óptica, é a Vicon que desenvolve estes sistemas desde 1984. Filmes como a *Guerra das Estrelas*, o *Gladiador* e *Titanic* recorreram à utilização do sistema *motion capture* da Vicon.

Alguns destes equipamentos conseguem capturar até 10 metros a grande velocidade. Estes estão divididos normalmente num núcleo (peça de hardware) que concentra o processamento, no equipamento de armazenamento (pode ser o computador), no sistema de comunicação e nas respectivas câmaras. As câmaras emitem infravermelhos de modo a que estes sejam reflectidos pelas marcas no actor identificando assim as várias posições. Nestes equipamentos as câmaras têm capacidade de capturar cerca de 120 marcas (reflectores), ou 400 marcas em modo *offline*. Quanto maior o número de marcas no actor, mais precisa será a captura. Quanto mais câmaras se utilizar, maior será a cobertura. A cada núcleo podem ser ligadas até 24 câmaras, cada câmara captura 250 fotogramas por segundo possibilitando a captura de movimentos rápidos. A qualidade da captura também varia consoante a resolução das câmaras. Nos equipamentos da Vicon as câmaras conseguem atingir resoluções na ordem dos 1000 x 1000 pixels. Também é possível capturar expressões faciais, podendo capturar expressões básicas com um número inferior de marcas na face, até capturar expressões faciais de alta qualidade com cerca de 100 marcas.

Configurar um sistema de captura de movimento óptico de raiz não é muito rápido, cerca de uma hora. As câmaras, os tripés e as luzes são colocados convenientemente a distâncias idênticas em volta da área de captura.

Depois das câmaras estarem todas a apontar para o centro da área de captura, é necessário calibrar as câmaras para dozes posições diferentes de um boneco de referência dentro da área de captura. Este processo ajuda a conhecer a localização relativa em relação às outras câmaras.

Este processo é feito apenas a primeira vez que se monta o sistema num determinado sítio, depois de estar configurado, é necessário apenas calibrar o sistema sempre que se volta a ligar. Demora cerca de dois minutos.

A resolução espacial deste sistema é de 0.0015 a 0.2% do FOV (Field of view) com uma precisão espacial de 0.004 a 0.5% do FOV e precisão temporal superior a 250 FPS, o espaço de trabalho é limitado ao raio de 1 a 4 metros. Multitrax, Optotrak, Vicon são alguns exemplos deste sistema capazes de funcionar com aplicações como os da Alias|Wavefront Maya, Discreet 3dStudio Max e Softimage XSI, Kaydara FilmBox.

### 2.2.5. Sistemas Acústicos

Este sistema é composto por um conjunto de emissores sonoros que são colocados nas principais articulações do actor. Três receptores sensíveis são posicionados no local da captura. Os transmissores são sequencialmente accionados para produzir um ruído característico, que será captado pelos receptores, calculando assim as suas posições no espaço.

O cálculo da posição de cada transmissor é feito da seguinte forma: utilizando como dados iniciais o tempo decorrido entre a emissão do ruído pelo transmissor e a sua recepção pelo receptor, e a velocidade do som no ambiente, consegue-se calcular a posição tridimensional de cada transmissor, é feita uma triangulação das distâncias deles em relação aos três receptores.

Um dos problemas deste método é a dificuldade de obter uma descrição correcta dos dados num instante desejado, devido ao carácter sequencial do disparo dos transmissores no corpo do actor. Além disso, os sistemas acústicos sofrem do mesmo tipo de problema existente em sistemas magnéticos: os incómodos cabos, que prejudicam a movimentação do actor. O número de transmissores que podem ser utilizados simultaneamente também é limitado, o que pode não fornecer uma descrição suficientemente correcta do movimento capturado.

Os sistemas acústicos estão sujeitos a problemas causados pela própria tecnologia: reflexões do som emitido pelos transmissores ou ruídos externos podem afectar o processo de captura e destruir os dados obtidos. Em compensação, este tipo de sistema não possui problemas de oclusão, típicos dos sistemas ópticos; e interferência por objectos metálicos como nos sistemas magnéticos.

A tecnologia de *motion capture* ainda está numa fase de desenvolvimento visto que os vários sistemas disponíveis ainda sofrem de muitos problemas de ordem técnica. No entanto, assistimos a uma rápida evolução destes sistemas devido à sua massificação, essencialmente nas indústrias da saúde, militar, espacial e entretenimento. Sendo estas indústrias detentoras de grandes capitais, o investimento nesta tecnologia faz com que evolua de uma forma exponencial, tornando este mercado extremamente apelativo para a investigação e para toda a indústria envolvente.

Actualmente adivinha-se que, com o ritmo da evolução desta tecnologia e com os preços dos equipamentos a tornarem-se mais acessíveis, esta irá atingir a massificação global.

### 2.3. *Face capture*

A análise e o reconhecimento de faces têm recebido intensa e crescente atenção da comunidade da Computação Gráfica, em parte devido às diversas aplicações, tais como interação homem computador, codificação baseada em modelos, vídeo-conferência, segurança e vigilância.

Uma tarefa importante que surge em diferentes problemas do reconhecimento de faces é a localização e segmentação de regiões de características faciais, tais como sobrancelhas, íris, lábios e narinas.

Para citar alguns exemplos de aplicação, a segmentação de tais características faciais é importante para o reconhecimento de faces baseado em características, aplicações médicas e análise de expressões faciais.

A extração das características faciais tem sido desde os anos 80 alvo de investigações especialmente ao nível da vídeo-conferência e compressão de vídeo. As suas aplicações são diversas como por exemplo, a segurança e a vigilância, a interação homem computador através de novas interfaces, a procura de pessoas desaparecidas, a computação afectiva, animação digital, e a vídeo-conferência.

Muitos sistemas de *motion capture* incluem o *face capture* como parte integrante do sistema, especialmente os sistemas ópticos, recorrendo a marcas na face para identificar as características faciais e as suas expressões. Mas existem métodos para o reconhecimento automático, é o caso da segmentação e o reconhecimento de regiões com características faciais utilizando homomorfismos entre grafos. Podemos entender regiões com características faciais como sendo sobrancelhas, íris, lábios e narinas. O método pode ser aplicado em imagens estáticas e até mesmo em sequências de vídeo.

Em primeiro lugar, uma imagem que contenha a fotografia do rosto de uma pessoa, a face é localizada. Este procedimento pode ser executado manualmente ou por meio da aplicação de uma técnica baseada na GWN (*Gabor Wavelet Networks*), a qual fornece uma localização aproximada das regiões de características faciais na imagem. A partir de então, a imagem é super-segmentada utilizando-se o algoritmo *watershed*. Esta imagem passa a ser representada por um grafo de atributos relacionais, o qual é rotulado convenientemente segundo um grafo modelo através de um homomorfismo. Este rótulo fornece-nos o reconhecimento das características faciais.

Este reconhecimento das características faciais pode ser posteriormente utilizado para encontrar uma correspondência com uma outra face, ou poderá servir como referência para animar uma outra.

### 3. Conclusão

O estudo desenvolvido permite tirar algumas conclusões sobre as várias tecnologias e a sua possível integração na proposta.

No que diz respeito à tecnologia de *motion capture*, é uma tecnologia que cada vez mais se afirma no mercado, especialmente com o aparecimento de muitas marcas a desenvolver novos sistemas e mais concorrentes. As suas utilizações começam a diversificar, e surgem novos mercados. No entanto é uma tecnologia que se baseia na captura do movimento real e especialmente para seres humanos, não sendo apropriada para a utilização em desenhos animados do estilo tradicional, outro grande entrave a esta tecnologia é o facto de ainda ser bastante dispendiosa.

Em relação ao *face capture*, e apesar de ser uma parte integrante do *motion capture* é uma tecnologia que tem tido abordagens diferentes, especialmente no que toca ao seu preço. Existem actualmente tecnologias de *face capture* que utilizam câmaras de vídeo normais para o registo de marcas ou dos traços faciais deixando todo o trabalho para o computador. A utilização do *face capture* neste projecto é ainda uma variável dependente da sua integração com os desenhos animados. Especialmente porque se pretende criar um sistema de desenhos animados bastante gráfico e não realista, o que de alguma forma não necessita de muito detalhe.

O mesmo se aplica para a tecnologia de reconhecimento da fala. Como se trata de um projecto virado para os desenhos animados gráficos e de tendência mais minimal, a utilização de um reconhecimento de fala numa plataforma destas seria puro desperdício de processamento. Para fazer o sincronismo labial, não é necessário identificar qual é a palavra que o actor está a dizer, tem é de ter uma resposta imediata, sendo de apostar numa linguagem não verbal, permitindo apenas o reconhecimento dos fonemas a partir dos volumes e das frequências do som. Desta forma poupa-se no processamento e consegue-se uma resposta em tempo real.

## **Bibliografia**

ALBERTO, Júnior Lucena (2002), *Arte da animação, técnica e estética através da história*, São Paulo, SENAC

BECK, Jerry (1998), Warner Bros., *Secrets et tradition de l'animation*, Paris, Dreamland

CHADWICK, John (1989), *Computer Graphics*, Nova Iorque, Siggraph

FURTINISS, Maureen (1998), *Art in Motion : Animation Aesthetics*, Sydney, John Libbey

GIRARD, Michael (1985), *Computer Graphics*, Nova Iorque, Siggraph

MORRISON, Mike (1994), *Becoming a Computer Animator*, Indianápolis, Howard W. Sams

RIVILIN, Robert (1986), *The Algorithmic Image: Graphic Vision of the Computer Age*, Microsoft Press

SPALTER, Anne (199), *The Computer in the Visual Arts*, Addison-Wesley

WILLIAM, Richard (2001), *The Animator's Survival Kit*, Londres, Faber and Faber

WILLIAM, T. Reeves (1981), *Computer Graphics*, Nova Iorque, Siggraph

IGARASHI, Takeo, HUGHES, T. John (2001), *Voice as Sound: Using Non-verbal Voice Input for Interactive*, Orlando, Siggraph / Eurographics

IGARASHI, Takeo, MOSCOVICH, Tomer, HUGHES, T. John (2005), *As-Rigid-As-Possible Shape Manipulation*, LA, Siggraph / Eurographics

IGARASHI, Takeo, MOSCOVICH, Tomer, HUGHES, T. John (2005), *Spatial Keyframing for performance-driven Animation*, Nova Iorque, Siggraph / Eurographics

KOCHANEK, Doris (1984), *Computer Graphics*, Nova Iorque, Siggraph

## **INTERNET**

<http://disney.go.com/corporate/>

<http://www.lucasarts.com/>

<http://www.pixar.com>

<http://www.vicon.com>

<http://www.dreamworks.com>

<http://www.realitymotioncapture.com>

<http://www.biovision.com>

<http://www.metamotion.com/>

<http://www.motionanalysis.com>

<http://pyros.com>

<http://www.merl.com/projects/puppets/>

[http://www.eurecom.fr/~image/Clonage/vc\\_mainpage.html](http://www.eurecom.fr/~image/Clonage/vc_mainpage.html)

<http://www.cyberware.com/>

<http://mambo.ucsc.edu/psl/fan.html>

<http://www.cs.cmu.edu/~cil/v-images.html>

<http://www.isel.ipl.pt/>

## Índice Remissivo

<b>A</b>	
Analogus .....	8
Animatton.....	8
ANN.....	7
Artificial Neural Networks .....	7
<b>C</b>	
CLIMA .....	7
<b>D</b>	
DCT2 .....	7
<b>E</b>	
ELITE.....	10
<b>F</b>	
Face Capture.....	14
FilmBox on-air .....	4
Flock of Birds .....	9
<b>G</b>	
Gabor Wavelet Networks .....	14
GWN.....	14
<b>H</b>	
HMM .....	5
<b>M</b>	
marionetas.....	3
Markov.....	5
Motion Analysis .....	7
Motion Capture.....	7
MultiTrax.....	10
<b>N</b>	
Nelly Nuts .....	3
<b>O</b>	
Optotrak.....	10
<b>R</b>	
Reconhecimento da fala.....	5
reconhecimento de faces.....	14
Redes neuronais.....	7
Resource Management .....	6
Rita Catita .....	3
<b>S</b>	
segmentação .....	14
Sistemas Acústicos.....	13
Sistemas Magnéticos.....	9
Sistemas Mecânicos .....	8
Sistemas Ópticos .....	10
<b>T</b>	
taxas de erro.....	6
Tecnologia .....	7
timecode.....	3
tracking confusion .....	11
Tv-animation .....	3
<b>U</b>	
Ultratrak .....	9
<b>V</b>	
Vicon.....	11
<b>W</b>	
watershed.....	14