

Pedro Campos

Sonificador de Imagens para Apoio a Utilizadores Invisuais



Departamento de Ciência de Computadores
Faculdade de Ciências da Universidade do Porto
2007

Pedro Campos

Sonificador de Imagens para Apoio a Utilizadores Invisuais



Luis Antunes

Departamento de Ciência de Computadores
Faculdade de Ciências da Universidade do Porto
2007

Pedro Campos

Sonificador de Imagens para Apoio a Utilizadores Invisuais



*Tese submetida na Faculdade de Ciências da
Universidade do Porto para obtenção do grau de Mestre
em Ciência de Computadores*

Departamento de Ciência de Computadores
Faculdade de Ciências da Universidade do Porto

2007

DCC/FCUP
BIBLIOTECA
Coloc. _____
N.º Reg. 1791
Departamento de Ciência de Computadores

Agradecimentos

Em primeiro lugar, gostaria de agradecer aos meus pais pela oportunidade e apoio que me deram para realizar este mestrado.

Agradeço também à Marta, a minha namorada, pela ajuda e apoio nos momentos que mais necessitei para a concretização da tese de mestrado.

Um muito obrigado ao professor Jorge Simão, pelo tempo despendido na orientação e realização da minha tese.

Um muito obrigado ao Rui Helder por se ter disponibilizado para avaliar, numa fase piloto, a usabilidade do software desenvolvido.

Agradeço também ao Pedro Silva que foi o interlocutor com a ACAPO (Associação dos Cegos e Amblíopes de Portugal). O meu agradecimento à ACAPO por terem tornado possível a interação com pessoas cegas e amblíopes afim de se avaliar a usabilidade da aplicação desenvolvida. Agradeço também individualmente aos voluntários cegos e amblíopes associados à ACAPO que testaram a usabilidade da aplicação, nomeadamente, ao António Pessoa, ao António Silva e à Margarida.

Resumo

Foi desenvolvido um sistema de software para conversão automática de imagens digitais e animações computacionais para audio (sonificação), com o intuito de apoiar utilizadores invisuais. O sistema usa técnicas de visão computacional para segmentar imagens e identificar objectos, e transforma informação posicional dos objectos e da sua fronteira para uma onda sonora sintetizada. Isso é feito usando dois tipos de referenciais cartesianos bidimensionais. Um tipo de referencial, localizado no centro-de-massa de cada objecto, é utilizado para sonificar a forma e a orientação dos objectos. Isso é feito através de um cursor que segue as arestas dos objectos detectados e convertendo a sua posição para som. O outro referencial está localizado num posição fixa e é utilizado para sonificar a posição do centro-de-massa dos objectos nesse sistema de coordenadas, que irá variar no tempo se o objecto se movimentar. Os sons gerados em associação com os dois referenciais são reproduzidos sequencialmente no tempo, de acordo com parâmetros ajustáveis. A conversão das posições (x, y) do cursor e do centro-de-massa para som é efectuada convertendo valores do eixo dos x 's para balanço do som, sendo os valores do eixo dos y 's convertidos para a frequência da onda sonora. A gama de frequências seleccionada é ajustável e diferente para os dois referenciais, permitindo ao utilizador distinguir o tipo de informação que está a ser sonificada (forma ou posição).

Índice

Índice de Tabelas	13
Índice de Figuras	16
1 Introdução	17
2 Trabalhos Relacionados	21
2.1 Sistema THE VOICE	21
2.2 Sistema SOUNDVIEW	22
2.3 Projecto EE/CS 107b	23
2.4 Sistema GUESS	23
2.4.1 O Interface GUESS	24
2.4.2 Soluções para localização de objectos no sistema GUESS	25
2.4.2.1 Localização usando as posições do quadro	25
2.4.2.2 Localização usando a grelha virtual	25
2.4.2.3 Localização usando o espectro 2D do som	26
3 Sonificação	27
3.1 Sonificação baseada em técnicas de computação visual: Generalidades .	27
3.1.1 Especificação formal	29

3.2	Um modelo para sonificação de imagens e animações	31
4	Detalhes de Implementação	43
4.1	Enumeração das funcionalidades da aplicação	43
4.2	Análise e desenho das fronteiras do objecto	45
4.2.1	Obtenção da fronteira dos objectos	45
4.2.2	Análise das fronteiras	47
4.2.3	Visualização de objectos	50
4.3	Tradução da forma e movimento do objecto para som	51
4.3.1	Cálculo de frequências e níveis de balanço para sonificação de objecto	51
4.3.2	Cálculo de frequências e níveis de balanço para sonificação do movimento	55
4.3.3	Reprodução acústica do objecto e da sua trajectória	57
4.4	Trajectórias pré-definidas para os objectos	59
4.4.1	Posição estática	59
4.4.2	Movimento rectilíneo uniforme	59
4.4.3	Movimento rectilíneo uniformemente acelerado	60
4.4.4	Movimento circular	60
4.4.5	Movimento oscilante	61
4.4.6	Movimento pendular	62
5	Avaliação da Usabilidade do Software	65
5.1	Primeiro protocolo experimental	66
5.2	Segundo protocolo experimental	67

6 Conclusões	75
Referências	79

Lista de Tabelas

5.1	Resultados do protocolo piloto experimental	68
5.2	Resultados do segundo protocolo experimental - 1º sujeito	70
5.3	Resultados do segundo protocolo experimental - 2º sujeito	71
5.4	Resultados do segundo protocolo experimental - 3º sujeito	72

Lista de Figuras

2.1	The vOICe.	21
2.2	Reprodução acústica de um triângulo recto.	24
2.3	Codificação de junções de linhas.	25
2.4	Grelha virtual de som.	26
2.5	Localização 2D do som.	26
3.1	Diagrama esquemático do processo de sonificação com múltiplos canais	31
3.2	Referenciais utilizados na sonificação	33
3.3	Filtro de contraste	34
3.4	Sonificação de um triângulo estático	35
3.5	Sonificação de um círculo estático	36
3.6	Sonificação do movimento rectilíneo uniformemente acelerado	38
3.7	Sonificação do movimento pendular	39
3.8	Série temporal das coordenadas do x e y do triângulo com movimento rectilíneo uniforme	40
3.9	Série temporal da sonificação do triângulo com movimento rectilíneo uniforme	40
3.10	Espaço-Fase visual do triângulo com movimento rectilíneo uniforme	41
3.11	Espaço-Fase audio do triângulo com movimento rectilíneo uniforme	41

4.1	Aplicação para sonificação de imagens e animações.	44
4.2	Diagrama de blocos da aplicação Movie2Sound.	45
4.3	Primeira fase do Movie2Sound.	45
4.4	Chain code	48
4.5	Sentidos de observação dos pixels vizinhos	49
4.6	Processo de construção de chain codes	50
4.7	Segunda fase do Movie2Sound.	51
4.8	Cálculo das constantes kh_{obj} e kv_{obj}	52
4.9	Frequências entre os pontos y_{min} e y_{max} para um objecto.	54
4.10	Balanço do volume entre os pontos x_{min} e x_{max} para um objecto.	54
4.11	Frequências entre os pontos inicial e final de uma trajectória.	57
4.12	Balanço do volume entre os pontos inicial e final de uma trajectória.	58
4.13	Posição estática.	59
4.14	Movimento rectilíneo uniforme.	60
4.15	Movimento rectilíneo uniformemente acelerado.	61
4.16	Movimento circular.	61
4.17	Movimento oscilante.	62
4.18	Movimento pendular.	62
6.1	Bola saltitante	76
6.2	Foguete espacial	76
6.3	Figuras icônicas	77

Capítulo 1

Introdução

Utilizadores de computadores cegos ou com deficiências visuais poderão demonstrar algum transtorno pela incapacidade de usar convenientemente sistemas computacionais multimédia, tal como acontece pelos os utilizadores em geral. Enquanto a informação presente numa frase de texto de um GUI pode ser facilmente substituída por mensagens de audio pré-gravadas para satisfazer as necessidades dos utilizadores cegos ou com deficiências visuais, a capacidade de transmitir informação visual contida em imagens e animações computacionais é muito mais difícil de se conseguir. Idealmente, deveriam-se utilizar soluções tecnológicas que permitam a utilizadores cegos "ver" imagens gráficas e/ou "ver" glifos animados, sem necessitar que os conteúdos multimédia sejam especialmente preparados para utilizadores cegos ou necessitar de hardware HCI caro e especial.

Um tipo de solução para a conversão automática de imagem para som (*sonificação*¹), é a captação de aspectos salientes nas imagens e usar um conjunto de convenções para converter informação visual para som. Diversas abordagens têm sido desenvolvidas pela comunidade científica seguindo estas linhas de orientação. Por exemplo, no sistema GUESS [KRS], a informação sobre a forma do objecto é traduzida para som através da conversão das coordenadas dos x 's para balanço do som, e as coordenadas dos y 's para frequência e/ou amplitude da onda sonora. O sistema GUESS foi projectado para permitir ao utilizador interagir com imagens digitais usando um quadro e uma caneta electrónica. A localização dos objectos e identificação da sua forma é feita

¹Obrigado a Mário Branco por sugerir a utilização da palavra sonificação para a conversão automática de informação visual para som.

através da exploração do quadro electrónico de tal forma que quando a caneta passa por cima de um objecto os contornos do objecto so automaticamente traduzidos para som. O sistema GUESS foi considerado como sendo uma boa solução para sonificação, e um avanço para algumas abordagens mais simples tais como converter os códigos de cor RGB ou grey-scale dos pixels das imagens para frequência e/ou padrões de som modulados em amplitude [Sue] [Mei] [vDD]. Contudo o sistema GUESS é utilizado somente para exploração de imagens e não pode ser facilmente utilizado para tradução de animações ou imagens em movimento, a não ser que o utilizador seja forçado a seguir activamente o movimento do objecto usando a caneta (o que é difícil).

Foi também realizado trabalho de investigação para incluir características de edição de animação em ferramentas especialmente projectadas para utilizadores cegos [KL]. No entanto, pouca pesquisa tem sido feita para criar soluções práticas para a sonificação automática de filmes ou animações gráficas para cegos. Trabalhando sobre a abordagem do sistema GUESS, desenvolvemos uma solução e uma aplicação protótipo para a conversão automática de imagens e animações computacionais para som adequado para ser utilizado como interface para cegos ou utilizadores com deficiências visuais. O sistema usa técnicas de visão computacional para segmentar imagens e identificar objectos, e transforma informação posicional dos objectos e da sua fronteira para uma onda sonora sintetizada. Isso é feito usando dois tipos de referenciais cartesianos bidimensionais. Um tipo de referencial, localizado no centro-de-massa de cada objecto, é utilizado para sonificar a forma e a orientação dos objectos. Isso é feito através de um cursor que segue as arestas dos objectos detectados e convertendo a sua posição para som. O outro referencial está localizado numa posição fixa e é utilizado para sonificar a posição do centro-de-massa dos objectos nesse sistema de coordenadas, que irá variar no tempo se o objecto se movimentar. Os sons gerados em associação com os dois referenciais são reproduzidos sequencialmente no tempo, de acordo com parâmetros ajustáveis. A conversão das posições (x, y) do cursor e do centro-de-massa para som é efectuada convertendo valores do eixo dos x 's para balanço do som, sendo os valores do eixo dos y 's convertidos para a frequência da onda sonora. A gama de frequências seleccionada é ajustável e diferente para os dois referenciais, permitindo ao utilizador distinguir o tipo de informação que está a ser sonificada (forma ou posição).

Para validar a utilidade da solução realizaram-se protocolos experimentais com utilizadores cegos, tendo sido retirados dados estatísticos sobre a identificação ou não de

estímulos visuais e respectivas sonificações pelos utilizadores. Os resultados encontrados mostram que para uma grande variedade de estímulos tanto em termos da forma dos objectos como do tipo dos seus movimentos, os utilizadores conseguem fazer uma identificação correcta e com alto nível de confiança.

Esta tese encontra-se subdividida em seis capítulos, procurando desta forma tornar mais fácil e imediata a sua leitura e compreensão.

Sendo assim,

- O primeiro capítulo contemplou a introdução, onde indicámos as razões que nos levaram a debruçar sobre o tema da sonificação, e onde apresentámos uma breve descrição do sistema implementado;
- No segundo capítulo, descreve-se trabalhos desenvolvidos e relacionados com a sonificação de imagens;
- O terceiro capítulo, apresenta algumas generalidades sobre técnicas de computação visual, do problema inerente à sonificação e de uma solução possível, bem como a especificação formal e o modelo experimental;
- No quarto capítulo apresenta-se as funcionalidades da aplicação e as diversas fases que compõem o processo de sonificação;
- No quinto capítulo ilustra-se os resultados obtidos de dois protocolos experimentais realizados com indivíduos amblíopes;
- Por último, no sexto capítulo, apresentar-se-á as conclusões que obtidas, bem como trabalhos futuros e aplicações práticas a desenvolver.

Capítulo 2

Trabalhos Relacionados

O tema da sonificação de imagens tem sido abordado usando uma pequena variedade de aproximações. Neste capítulo apresentamos uma descrição sumária das diferentes aproximações seguidas por outros investigadores. Nomeadamente, na secção 2.1 apresentamos o sistema THE VOICE [Mei]. Na secção 2.2 descrevemos o sistema SOUNDVIEW [vDD]. Na secção 2.3 apresentamos o projecto EE/CS 107b [Sue] e por último na secção 2.4 descrevemos o sistema GUESS [KRS].

2.1 Sistema The vOICe

A tecnologia de visão THE VOICE para utilizadores invisuais [Mei], oferece a possibilidade de sonificar imagens digitais fazendo deslocar uma barra vertical ao longo do eixo horizontal da imagem da esquerda para a direita, funcionando assim o sistema à semelhança de um *scanner*. Para cada posição da barra o sistema identifica todas as zonas pintadas da imagem e produz uma onda sonora multi-frequência associando a pixels pintados em posições superiores a frequências altas e pixels em posições inferiores a frequências baixas. A informação sobre a posição da barra é transmitida modulando o balanço da onda sonora.



Figura 2.1: The vOICe.

Esta tecnologia permite a pessoas invisuais transportar um PC portátil e uma webcam e ouvir as imagens do ambiente que as rodeia através dos auscultadores. Por exemplo, um ponto de luz brilhante produz um beep curto. Se o ponto de luz estiver situado à sua esquerda, ouve-se o beep no auscultador da esquerda e se estiver à sua direita, ouve-se o beep no auscultador da direita. Se o ponto de luz se mover para cima, ouvir-se-á a frequência do beep a subir e se mover para baixo, a frequência do beep também desce.

O número de beeps é proporcional ao número de pontos de luz que existe na imagem. Uma linha brilhante horizontal dá origem a um longo som, porque os pontos que compõem esta linha correspondem a uma concatenação dos beeps no tempo, parecendo ouvir-se como um som contínuo. Uma linha vertical é uma pilha de pontos, ouvindo-se todos ao mesmo tempo mas todos com frequências diferentes visto que se encontram a alturas diferentes. No conjunto ouvir-se-á um breve ruído.

Naturalmente quanto mais complexa for a imagem, mais complexo será o som. Contudo demora sempre um segundo para se ouvir o conteúdo de uma imagem inteira.

Como o sistema está preparado para funcionar com sequências de imagens em *real time*, o sistema repete o processo de varrimento continuamente, sendo o número de varrimentos igual ao número de segundos em que o sistema esteve operacional.

2.2 Sistema SoundView

SOUNDVIEW é um sistema experimental que permite a exploração da cor de uma imagem através do toque e da audição [vDD]. A imagem é convertida para uma superfície virtual com uma textura sonora fazendo corresponder cores dos pixels a sons variados. O utilizador explora a imagem movendo um dispositivo apontador por cima da imagem (*light-pen*), que age como uma agulha virtual de gramafone que produz som dependente do movimento e da área explorada. Para realizar a sonificação das cores cria-se um isomorfismo entre o espaço de cores e o espaço dos sons produzidos.

A cor de um pixel é representada no espaço HSB (hue, saturation, brightness). A transformação entre som e cor envolve uma série de filtros e heurísticas para modulação de frequências tendo a solução final sido desenhada numa base empírica, permitindo ainda que os detalhes da sonificação sejam ajustáveis pelo um conjunto de parâmetros

de controle. De uma forma geral devido ao mapeamento isométrico, cores semelhantes são mapeadas para sons semelhantes.

O sistema SOUNDVIEW foi implementado em JASS (Java Audio Synthesis System), usando um painel gráfico da Wacom como dispositivo de entrada. Devido ao ruído existente na posição do apontador é utilizado um filtro lowpass para suavizar a movimentação do apontador, que é utilizado a uma taxa fixa de $44100\text{Hz}/64 = 689\text{Hz}$. A cada instante em que a posição da caneta é lida, 64 amostras de um buffer de audio, são processadas através de um filtro gráfico após ter computado e ajustado as variáveis do filtro de acordo com a cor do pixel da velocidade do apontador. A fim de invocar a ilusão ao toque, é crucial uma baixa latência do audio. Isto é conseguido usando bibliotecas nativas para aceder a recursos específicos de audio no sistema operativo. O sistema actualmente funciona em Windows, LINUX e MAC OS/X.

2.3 Projecto EE/CS 107b

Foi desenvolvido um projecto em Matlab (publicado apenas na WEB), que consiste num módulo capaz de gerar impressões sonoras a partir de imagens [Sue]. À semelhança do sistema SOUNDVIEW, baseia-se na conversão da informação de cor dos pixels da imagem para uma onda sonora. Faz uso de um conjunto de transformações elaboradas para determinar as frequências do som gerado. Varrimentos da imagem na direcção horizontal produzem modulação de som num dos auscultadores/colunas (esquerda ou direita), e varrimentos da imagem na direcção vertical produzem modulação do som no outro auscultador/coluna. Como o objectivo do sistema não é permitir a inferência das formas contidas nas imagens, mas apenas causar uma impressão sensorial de carácter artístico no utilizador, torna-se difícil ou mesmo impossível avaliar a qualidade da solução proposta. Por esse motivo, não se apresentam mais detalhes sobre esta solução.

2.4 Sistema GUESS

O sistema GUESS permite aos utilizadores utilizar uma caneta electrónica para obter informação topológica de uma dada imagem [KRS]. Com essa caneta electrónica, o utilizador cego interage com um quadro gráfico para explorar uma imagem (e.g,

composta de formas geométricas simples). Durante essa exploração, o utilizador pode obter dois tipos de informação auditiva. Nomeadamente, informação sobre os contornos das formas, e disposição espacial de cada forma relativamente à imagem inteira.

2.4.1 O Interface GUESS

Na aproximação do sistema GUESS, o utilizador pode movimentar livremente a caneta eletrónica sobre o quadro (*pad*) de imagens afim de explorar a imagem e identificar a posição dos objectos nela presentes. São usadas diferentes soluções para ajudar o utilizador a encontrar as figuras contidas na imagem (descritas em baixo).

Quando uma figura presente numa imagem é encontrada, ela é traduzida automaticamente para som, percorrendo a fronteira do objecto e gerando uma onda sonora contínua que é modulada ao longo do movimento feito pela fronteira. Especificamente deslocamentos horizontais correspondem a modulação de balanço, e a movimentos verticais faz-se corresponder variações de frequência. Por exemplo, se a forma traduzida é um triângulo recto, o utilizador ouve um tom descendo verticalmente no canal direito. De seguida, o som move-se horizontalmente desde do fundo à direita para o canal ao fundo à esquerda. Finalmente, ascende da inferior esquerda para o canal superior direito, alcançando a posição inicial (ver figura 2.2).

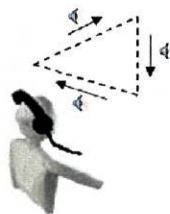


Figura 2.2: Reprodução acústica de um triângulo recto.

Para facilitar a identificação das imagens a partir do som, foi utilizado o som do alarme (*beep*) para codificar vértices ou junções entre linhas (ver figura 2.3). A solução de sonificação usada nesta tese é (parcialmente) inspirada na solução implementada pelo sistema GUESS mas no nosso trabalho demos também atenção ao aspecto de sonificação de movimento.

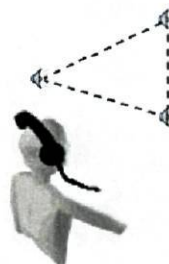


Figura 2.3: Codificação de junções de linhas.

2.4.2 Soluções para localização de objectos no sistema GUESS

Em baixo descrevemos as três técnicas utilizadas no sistema GUESS para auxiliar o utilizador a localizar os objectos gráficos presentes nas imagens.

2.4.2.1 Localização usando as posições do quadro

Na solução mais simples para identificação de posição, assume-se que o utilizador é capaz de inferir a posição da caneta dentro do quadro baseado apenas na noção que tem da posição da sua mão (usando os mecanismos de localização espacial dos humanos). Para facilitar o processo, coloca-se o quadro electrónico numa posição fixa e frontal em relação ao utilizador. Quando uma certa área contendo uma forma geométrica é alcançado, o algoritmo de sonificação descrito em cima é activado.

2.4.2.2 Localização usando a grelha virtual

Na segunda técnica, foi utilizado um modelo baseado numa grelha para aumentar a precisão de exploração do painel. Foi traçado uma grelha virtual 3×3 no quadro, consistindo em 9 regiões não sobrepostas e dispostas como um *Keypad* do telefone (“1” é a célula superior esquerda da grelha e “9” a célula inferior direita - ver figura 2.4). Foi atribuído aos eixos verticais e horizontais o som de dois instrumentos musicais diferentes. A fim de comunicar a localização de qualquer célula individual na grelha são reproduzidas duas notas consecutivamente sempre que se cruza a fronteira e que se entra numa célula. Nomeadamente, para discriminar as posições das células ao longo do eixo horizontal são atribuídos diferentes tons de *clarinete* a cada posição, e para discriminar as posições das células ao longo do eixo vertical são atribuídos diferentes

sons de *vibrofone*. A frequência das duas notas de *clarinete* são aumentadas à medida em que a caneta electrónica se move da esquerda para a direita; similarmente, as notas de *vibrofone* diminuem à medida que a caneta se move do topo para o fundo. Além disso, foi atribuído um som de um tambor à célula central (número “5”) para mais fácil identificação (ver figura 2.4).

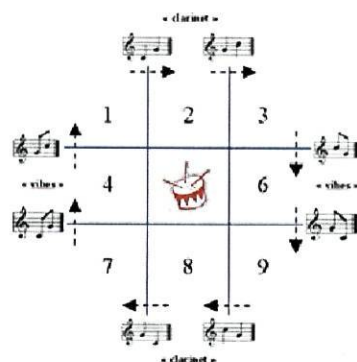


Figura 2.4: Grelha virtual de som.

2.4.2.3 Localização usando o espectro 2D do som

Na terceira e última técnica de exploração e localização é modulado o volume do som tornando-se mais elevado quanto mais próximo a caneta estiver de um objecto. O objecto sonificado é sempre aquele cujo o centro esteja mais próximo da posição actual da caneta. O volume de sonificação é máximo quando a caneta se encontra no centro de uma figura geométrica (ver figura 2.5).

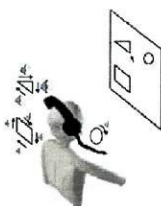


Figura 2.5: Localização 2D do som.

Capítulo 3

Sonificação

A sonificação de imagens envolve a conversão das dimensões visuais das entidades presentes numa imagem para dimensões que caracterizam ondas audio¹. É possível conceber uma variedade de soluções concretas para o problema da sonificação. Por isso, neste capítulo discutiremos os vários aspectos que podem estar envolvidos em processo de sonificação e apresentamos uma descrição sumária do algoritmo de sonificação concreto usado nesta tese. Nomeadamente, na secção 3.1 apresentamos uma descrição genérica dos processos de sonificação de imagens e animações, incluindo uma possível caracterização formal desses processos. Na secção 3.2 apresentamos um modelo concreto para sonificação de imagens e animações.

3.1 Sonificação baseada em técnicas de computação visual: Generalidades

A sonificação de imagens envolve a conversão das dimensões visuais dos objectos ou texturas presentes na imagem para dimensões do som. As dimensões visuais podem ser consideradas a vários níveis de análise e complexidade: ao nível dos objectos ou regiões da imagem podem ser considerados aspectos/dimensões, tais como localização, dimensão, orientação e forma; ao nível do pixel, podem-se considerar aspectos como a localização e o valor dos pixels (código de cor e/ou intensidade). A análise ao nível

¹Este capítulo é baseado no artigo “Sonification of Form and Movement for Visual-Impaired Users”, DSAI 2007 [JSaS].

do objecto pode também incluir o reconhecimento de classes de objectos presentes na imagem, as quais podem ser sonificadas directamente para representações linguísticas (e.g. o reconhecimento que um tipo particular de objecto presente numa imagem, tal como uma cadeira, ou uma específica espécie de animal, tal como um cão, pode despoletar a produção do som das palavras "cadeira" e "cão"). Outros níveis de análise poderão ser considerados (e.g. *features* geométricas, entre o nível do pixel e do objecto, ou ao nível da cena completa).

As dimensões do som podem também ser consideradas (pelo menos) a dois níveis de complexidade: características simples do som, que inclui tempo, *pitch*/frequência e amplitude/intensidade; características mais complexas do som, incluem o timbre (ou forma de onda), envelope, entre outras. No contexto de um canal de comunicação entre um emissor ou uma entidade produtora de som e um receptor, a entidade que escuta o som, podem ser consideradas dimensões adicionais. Nomeadamente, a posição espacial e/ou orientação da fonte em relação ao ouvinte, ou em relação a um referencial externo.

Quando estão envolvidas animações (sequências de imagens ou glifos animados), outras dimensões dos conteúdos multimédia podem ser considerados, tais como: alteração da posição (velocidade) e/ou aceleração de objectos, alterações da forma do objecto, fluxo óptico global, entre outras.

Qualquer solução para o problema de sonificação envolve os três seguintes pontos:

- Seleccionar quais as dimensões da informação visual que serão traduzidas para o utilizador.
- Seleccionar quais as dimensões do som que irão ser moduladas/sintetizadas para produzir variados estímulos auditivos.
- Determinar como é que a informação seleccionada nas dimensões visuais é transformada para as dimensões de audio consideradas.

Um problema básico da sonificação é que as imagens têm muita informação e a capacidade natural dos humanos para processar informação auditiva é normalmente mais baixa do que para processar informação visual. Assim, passar uma grande quantidade de informação visual para o utilizador numa forma pouco estruturada (e.g.

3.1. SONIFICAÇÃO BASEADA EM TÉCNICAS DE COMPUTAÇÃO VISUAL: GENERALII

produzindo sons para todos os pixels da imagem), é provável que crie no utilizador uma sobrecarga auditiva/cognitiva. Para tornar a carga cognitiva aceitável, somente os aspectos chave da informação visual deverão ser sonificados.

Uma complicação relacionada com a sonificação é o facto de o número de dimensões de informação visual que se pretende traduzir para o utilizador, possa ser muito maior do que o número de dimensões do som que podemos ou que é fisicamente possível de manipular. Assim, pode ser necessário usar compromissos no processo de conversão de imagens para som. Por exemplo, algumas dimensões da informação visual podem ser descartadas, dando prioridade a dimensões que sejam consideradas mais importantes (e.g. ignorar-se a informação de cor ou a textura dos objectos). Alternativamente, várias ondas de audio podem ser sequenciadas no tempo, fazendo as dimensões visuais de alta prioridade ocuparem mais tempo do que as dimensões visuais de baixa prioridade.

3.1.1 Especificação formal

Formalmente, o problema da sonificação de imagens pode ser descrito da seguinte forma: dado uma imagem img , define-se um conjunto de operadores visuais V_c da forma:

$$V_c(img) \rightarrow v_c^*$$

onde c é um index que discrimina entre um conjunto de operadores visuais (representando diferentes canais), e cujo output v_c^* é uma sequência contendo em cada posição i um vector $v^i \equiv [f_1^i, \dots, f_j^i, \dots, f_n^i]$. f_j^i contém algum tipo de informação/característica sobre a imagem a ser sonificada (e.g. valor das coordenadas x e y de um ponto da imagem, código de cores, identidades de objectos reconhecidos, etc.).

Cada vector $v^i \in v_c^*$, é transformado para uma forma sonora $s^i \equiv [g_1^i, \dots, g_k^i, \dots, g_m^i]$. g_k^i codifica um qualquer atributo da onda de som a ser produzida (e.g. balanço, *pitch* ou frequência, volume, identificação de uma amostra de som, descrição da onda de som, etc.). Um canal visual v_c^* é, portanto, transformado para uma sequência de vectores de atributos, que designaremos por s_c^* , representando uma onda sonora (possivelmente complexa). Representamos essa transformação pelo operador de sonificação $S_c(v_c^*)$ definido como:

$$S_c(v_c^*) \rightarrow s_c^*$$

Cada vector $s^i \in s_c^*$ será usado para gerar uma onda audio de acordo com os atributos $g_k^i \in s^i$. Essa onda deve reflectir a informação dos vectores g_k^i pela ordem que aparecem em s_c^* .

Formalmente, podemos definir relações de ordem total \prec_c e \prec'_c para cada canal visual e sonoro, e escrever $s^{j_1} \prec'_c s^{j_2}$, se s^{j_1} aparece primeiro do que s^{j_2} em s_c^* , e escrever $v^{j_1} \prec_c v^{j_2}$ se v^{j_1} aparecer primeiro do que v^{j_2} em v_c^* .

A ordem dos vectores $v^i \in v_c^*$ deve ser preservada na sua correspondente forma do som $s^i \in s_c^*$. Isto é, se v^{j_1} aparecer primeiro em v_c^* do que em v_{j_2} , então a correspondente forma do som s^{j_1} deve também aparecer em s_c^* primeiro do que em s^{j_2} , e consequentemente ser reproduzida em primeiro no tempo. Formalmente, se:

$$v^{j_1} \prec_c v^{j_2} \Rightarrow s^{j_1} \prec'_c s^{j_2}$$

Quando são utilizados múltiplos canais visuais e sonoros, deve ser tomada a decisão de quando e qual canal de som deverá ser reproduzido a um dado instante. Formalmente, isto pode ser descrito através da introdução do operador selector do som M definido como:

$$M(\{s_c^*\}; t) \rightarrow A(t) \subset \{s_c^*\}$$

tomando como input o conjunto de todas as sequências de informação sonora (um por cada canal), e dando como output a identidade dos canais de som a reproduzir em cada instante do tempo t . Uma solução concreta para $M(\{s_c^*\}; t)$ é reproduzir os diversos canais simultaneamente. Outra solução é reproduzir em sequência. É útil definir este operador para sonificação de múltiplos canais, por exemplo, quando um canal é usado para sonificar a localização de um objecto e outro canal para sonificar a sua posição.

Para generalizar a especificação acima afim de se sonificar animações gráficas ou vídeos, ao invés de simples imagens estáticas, modifica-se o input dos operadores visuais V_c para sequências de imagens img^* . Assim, obtemos a definição:

$$V_c(img^*) \rightarrow v_c^*$$

Para alguns formatos de codificação de animações, poderá ser mais prático tratar o vídeo como um conjunto de canais atribuídos a cada objecto presente nas cenas.

A figura 3.1 mostra a representação esquemática do processo de sonificação como foi especificado nesta secção.

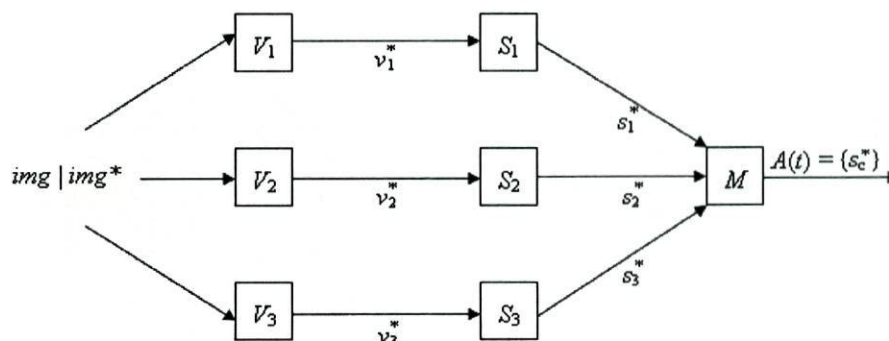


Figura 3.1: Diagrama esquemático do processo de sonificação com múltiplos canais: V_c é um operador visual que gera vectores de informação visual v_c^* a partir de imagens img ou animações img^* ; S_c é um operador de sonificação que gera representações de ondas audio s_c^* ; as ondas audio s_c^* são entregues a um selector de múltiplos canais M , que especifica qual o subconjunto de canais de som $A(t) = \{s_c^*\}$ deve ser reproduzido em cada instante do tempo t .

3.2 Um modelo para sonificação de imagens e animações

No nosso modelo experimental de sonificação proposto produz-se informação sobre a posição e forma de objectos. Isto é conseguido através de uma transformação entre os pontos (x, y) de algum referencial e sistema de coordenadas para uma onda audio com características que depende do tuplo (x, y) . A transformação concreta que utilizamos consiste em transformar os valores das coordenadas do eixo dos x 's para balanço (diferença no volume do som entre o aparelho de reprodução audio para o lado direito e para o lado esquerdo, tais como auscultadores ou colunas de som). Mais especificamente, valores positivos de x produzem sons com volume maior no

aparelho de audição da direita e valores negativos de x produzem sons com volume maior no aparelho de audição da esquerda. O total da intensidade do som é mantida constante². A motivação para este design é considerar que o utilizador se localiza/situa na origem do referencial considerado, e que o som reproduzido reflecta a *sensação* que um utilizador teria se a fonte do som se situasse na posição (x, y) . Os valores das coordenadas do eixo dos y 's são convertidos para *pitch*/frequência do som. Valores baixos de y são convertidos para baixas frequências (dentro de uma gama de frequências escolhida) e valores altos de y convertidos para altas frequências. Esta solução é motivada pela tendência psicológica dos humanos em associar altas frequências do som com localizações altas no espaço e baixas frequências do som a posições perto do chão. Podemos resumir este desenho/algoritmo definindo canais visuais e de audio da seguinte forma:

- Canais visuais — $v_c^* = [x, y]^*$, $c \in \{0, 1\}$, onde x, y são pontos de coordenadas no referencial associado com o canal.
- Canais de som — $s_c^* = [b, f]^*$, $c \in \{0, 1\}$, onde b é o balanço do som ($b > 0$, indica amplitudes de som maiores na coluna da direita e $b < 0$ indica amplitudes de som maiores na coluna da esquerda).

A transformação da informação do canal visual para o canal do som:

$$s([x, y]) = [b(x), f(y; [f_{\min}, f_{\max}])]$$

onde $b(x)$ representa a transformação dos valores das coordenadas do x para balanço e $f(y; [f_{\min}, f_{\max}])$ representa a transformação dos valores das coordenadas y para valores de frequência de uma gama de frequências seleccionada $[f_{\min}, f_{\max}]$, fica definida por:

$$\begin{aligned} b(x) &\propto x \\ f(y; [f_{\min}, f_{\max}]) &\propto y \end{aligned}$$

Dois tipos de referenciais são utilizados na solução proposta, correspondendo a dois tipos de canais imagem-audio. Um tipo de referencial fica situado no centro-de-massa

²Estudos preliminares de usabilidade provaram que esta solução é mais cognitivamente adequada para os utilizadores do que soluções onde o som seja mais alto perto de $x = 0$ ou o mais afastado de $x = 0$.

do objecto que está a ser considerado e é utilizado para reproduzir a informação sobre a forma do objecto (assim, se múltiplos objectos estão presentes no ecrã, existirá um referencial e canal para cada objecto). O outro referencial considerado fica situado num ponto fixo da imagem (nomeadamente, no pixel central do ecrã). Este é utilizado para reproduzir informação sobre a localização do centro-de-massa do objecto e do seu movimento. Este referencial é único e independente do número de objectos que estão a ser analisados numa imagem ou animação. A figura 3.2 esquematiza os dois tipos de referenciais de coordenadas utilizados. Os eixos dos referenciais estão sempre alinhados com as direcções horizontal e vertical do ecrã/imagem.

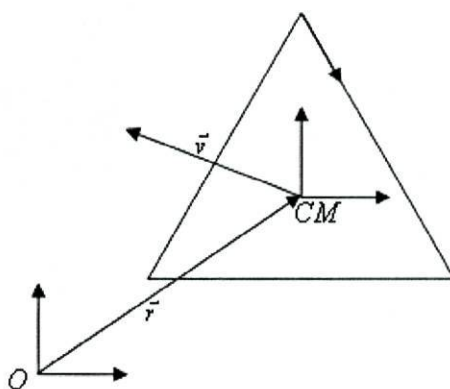


Figura 3.2: Dois tipos de referenciais utilizados na sonificação: examinando à volta das arestas do objecto gera os valores das coordenadas (x, y) no referencial do centro-de-massa do objecto — utilizado para reproduzir a informação sobre a forma (forma e orientação); alterações nas coordenadas do centro-de-massa do objecto (quando animado) produz alterações na onda de som — utilizado para reproduzir informação sobre a posição do objecto em relação a um referencial fixo com origem em O .

Dado um objecto k presente numa imagem, fazemos “passar” um cursor $p_k \equiv (x_k, y_k)$ sobre a fronteira do objecto, de tal forma que as coordenadas (x_k, y_k) produzem uma sequência de informação visual v_k^* que será sonificada como foi explicado anteriormente. Os valores das coordenadas (x_k, y_k) são definidos em relação ao centro-de-massa do objecto.

Em preparação prévia para obter a informação sobre a localização do cursor da fronteira do objecto, utilizamos primeiro um operador de contraste de imagem. Em particular, utilizamos um operador de contraste de gradiente de primeira ordem:

$F(img) \rightarrow img' : \quad img'(x, y) = |\nabla img(x, y)| > \theta$, onde θ é o valor fixo de threshold de contraste. A figura 3.3 mostra um exemplo de uma imagem (com uma forma triangular) e a imagem de contraste resultante.

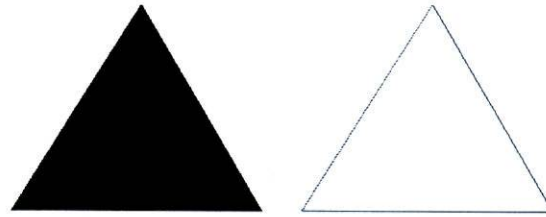


Figura 3.3: Aplicação standard de um filtro de contraste numa imagem para obter as arestas do objecto. **esquerda)** imagem fonte; **direita)** imagem filtrada.

Depois de se obter o contraste da imagem e as arestas do objecto, a imagem resultante é pesquisada pixel a pixel para identificar as coordenadas de um ponto pertencente às aresta de um objecto sendo utilizado como coordenada inicial para o cursor. Para gerar a sequência de informação visual tendo em conta a forma do objecto, usa-se as coordenadas do cursor que é movimentado em cada instante para o pixel adjacente à posição actual com maior valor de contraste.

Como ilustração do método, as figuras 3.4topo e 3.5topo mostram, respectivamente, a série temporal para a sonificação de um triângulo e de um círculo, mostrando os valores das coordenadas do cursor ao longo do tempo, e a frequência e balanço (mais o volume da coluna da esquerda e da direita em separado) do som gerado. As figuras 3.4baixo e 3.5baixo ilustram a representação da sonificação como uma transformação de um espaço geométrico bidimensional para o espaço-fase de modulação do som. Note-se que estes dois espaços (imagem e audio) são isomorficos no sentido matemático/topológico, visto que existe uma preservação estrutural da topologia dos espaços com transformações em ambos os sentidos. Significa isto, que a localização do cursor e a forma do objecto pode ser inferida a partir do som. Supõe-se que esta operação será implementada pelo cérebro do utilizador cego enquanto ouve o som gerado (segundo um processo neuronal-fisiológico não conhecido).

Para a geração da informação sobre o movimento do objecto, geramos um vector com a localização visual da posição do centro-de-massa do objecto e usamos o mesmo método de sonificação como acima descrito para a forma do objecto mas utiliza-se uma gama de frequências diferente (mais baixa ou mais alta). Como exemplo, ilustrámos

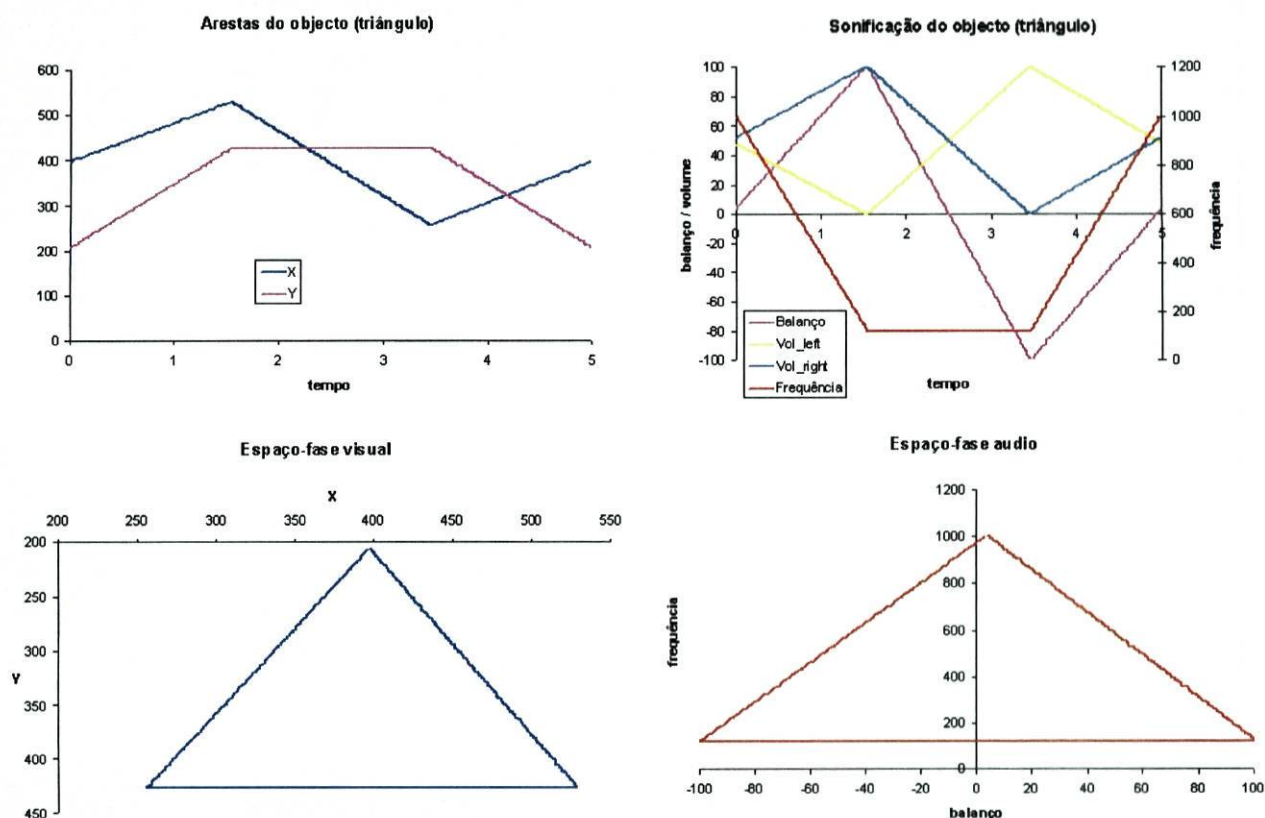


Figura 3.4: Sonificação de um triângulo estático: **topo-esquerda**) Série temporal das coordenadas do x e y quando faz a pesquisa das arestas/fronteira do triângulo; **topo-direita**) Série temporal da sonificação do triângulo: transformação dos valores das coordenadas do y das arestas para a frequência do som e transformação dos valores das coordenadas do x para balanço numa escala de $[-100, 100]$, com -100 representa volume da coluna esquerda no máximo e a da direita no mínimo e 100 o inverso (volumes em separado da coluna esquerda e da direita também visualizados); **baixo-esquerda**) representação das arestas do triângulo como movimento visual num estado espacial (x, y); **baixo-direita**) representação das arestas do triângulo como movimento num estado espacial do som (*frequência, balanço*).

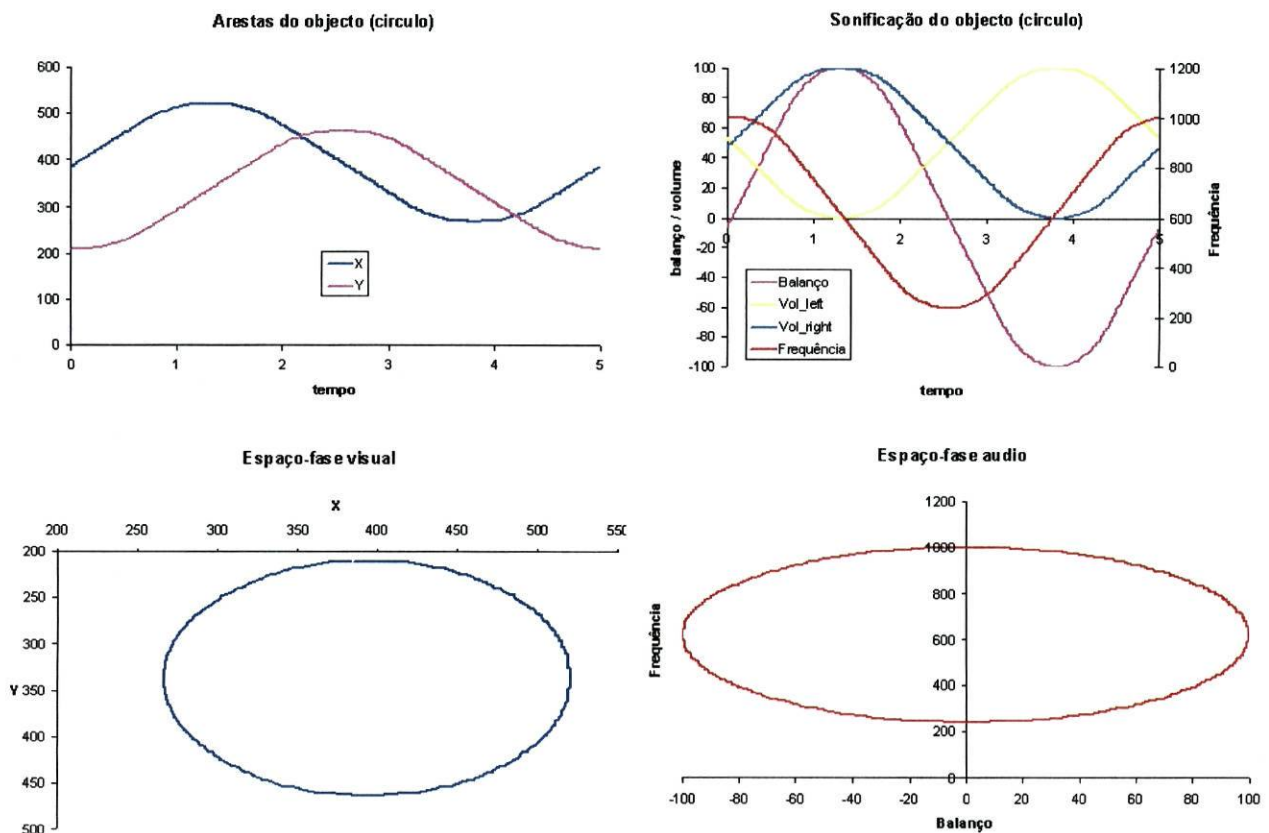


Figura 3.5: Sonificação de um círculo estático: **topo-esquerda**) Série temporal das coordenadas do x e y quando faz a pesquisa das arestas/fronteira do círculo; **topo-direita**) Série temporal da sonificação do círculo: transformação dos valores das coordenadas do y das arestas para a frequência do som e transformação dos valores das coordenadas do x para balanço numa escala de $[-100, 100]$, com -100 representa volume da coluna esquerda no máximo e a da direita no mínimo e 100 o inverso (volumes em separado da coluna esquerda e da direita também visualizados); **baixo-esquerda**) representação da fronteira do círculo como movimento visual num estado espacial (x, y) ; **baixo-direita**) representação da fronteira do círculo como movimento num estado espacial do som (*frequência, balanço*).

respectivamente, o movimento rectilíneo uniformemente acelerado (figura 3.6) e o movimento pendular (figura 3.7).

Quando combinámos um objecto com um tipo de movimento, o objecto é sonificado em primeiro lugar e de seguida o tipo de movimento escolhido. As figuras 3.8 à 3.11 representam um triângulo com movimento rectilíneo uniforme e ilustram o método acima descrito.

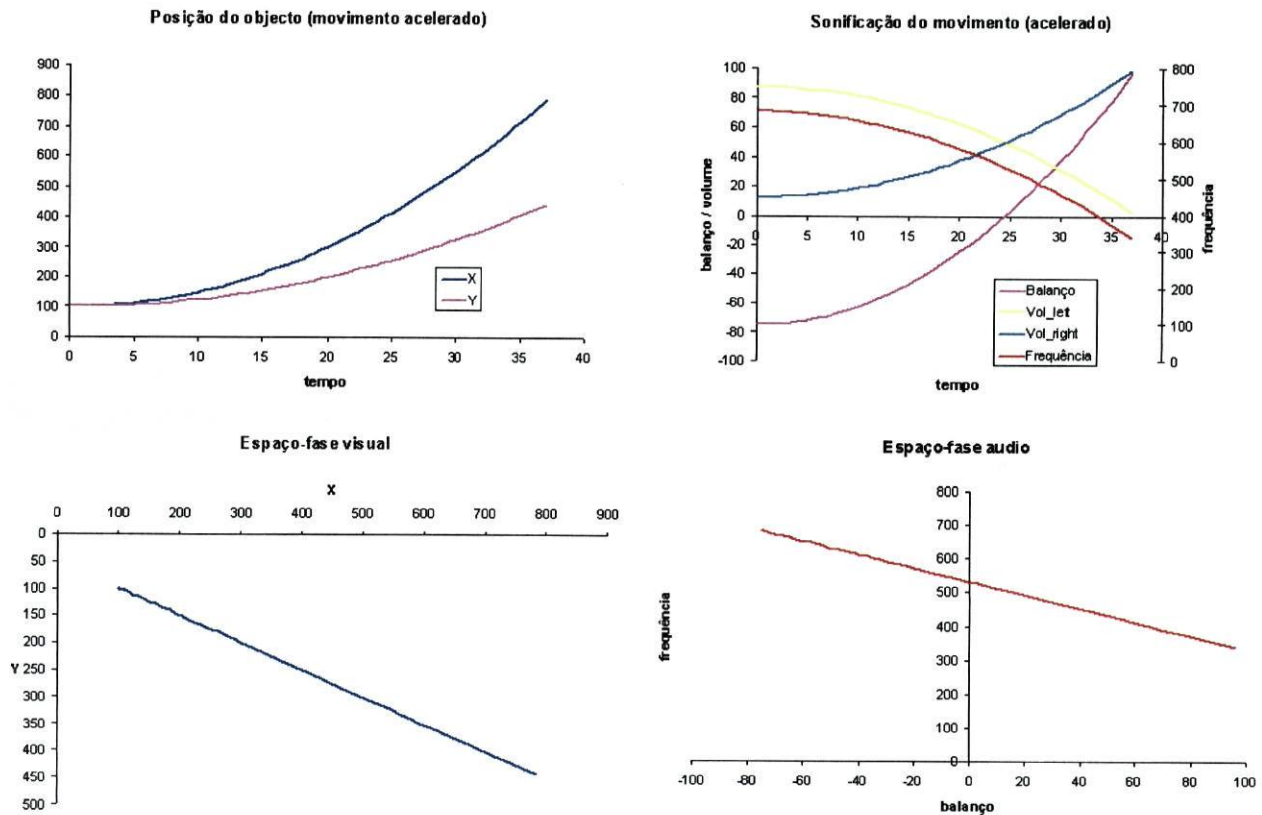


Figura 3.6: Sonificação do movimento rectilíneo uniformemente acelerado: **topo-esquerda**) Série temporal das coordenadas do x e y da posição do centro-de-massa do objecto; **topo-direita**) Série temporal da sonificação do movimento rectilíneo uniformemente acelerado: transformação dos valores das coordenadas do y da trajectória para a frequência do som e transformação dos valores das coordenadas do x para balanço numa escala de $[-100, 100]$, com -100 representa volume da coluna esquerda no máximo e a da direita no mínimo e 100 o inverso (volumes em separado da coluna esquerda e da direita também visualizados); **baixo-esquerda**) representação da posição do centro-de-massa do objecto como movimento visual num estado espacial (x, y) ; **baixo-direita**) representação da posição do centro-de-massa do objecto como movimento num estado espacial do som (*frequência, balanço*).

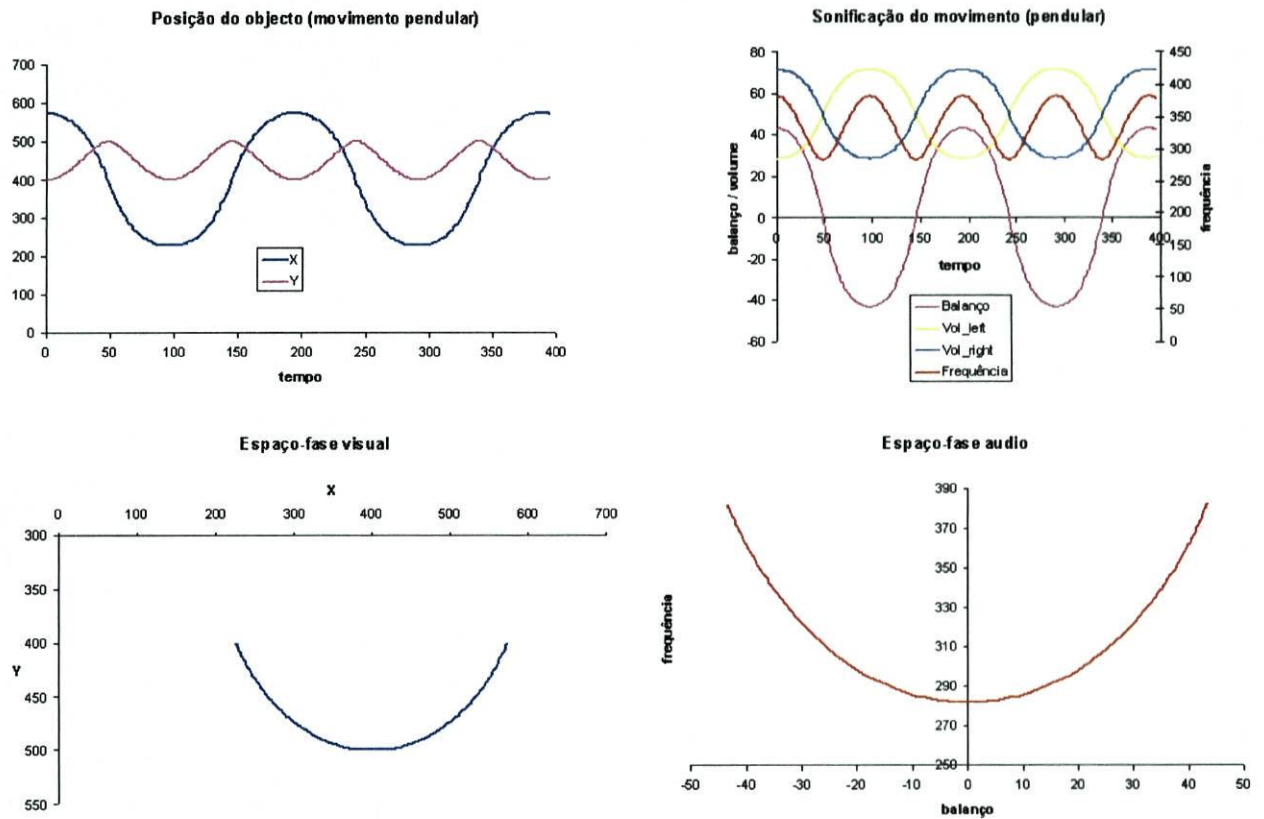


Figura 3.7: Sonificação do movimento pendular: **topo-esquerda**) Série temporal das coordenadas do x e y da posição do centro-de-massa do objecto; **topo-direita**) Série temporal da sonificação do movimento pendular: transformação dos valores das coordenadas do y da trajetória para a frequência do som e transformação dos valores das coordenadas do x para balanço numa escala de $[-100, 100]$, com -100 representa volume da coluna esquerda no máximo e a da direita no mínimo e 100 o inverso (volumes em separado da coluna esquerda e da direita também visualizados); **baixo-esquerda**) representação da posição do centro-de-massa do objecto como movimento visual num estado espacial (x, y) ; **baixo-direita**) representação da posição do centro-de-massa do objecto como movimento num estado espacial do som $(frequência, balanço)$.

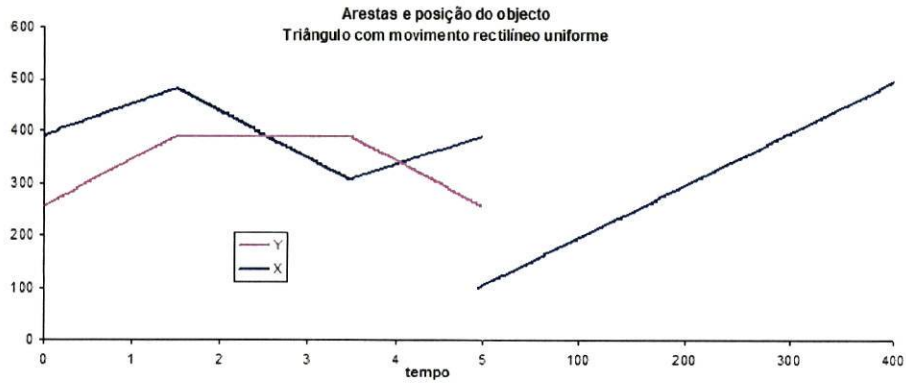


Figura 3.8: Sonificação do triângulo com movimento retilíneo uniforme: Série temporal das coordenadas do x e y quando faz a pesquisa das arestas/fronteira do triângulo e da posição do centro-de-massa do mesmo.

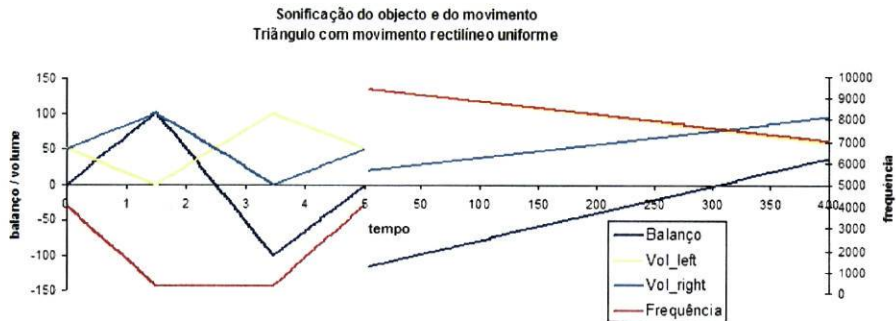


Figura 3.9: Série temporal da sonificação do triângulo com movimento retilíneo uniforme: transformação dos valores das coordenadas do y das arestas e da trajectória para a frequência do som e transformação dos valores das coordenadas do x para balanço numa escala de $[-100, 100]$, com -100 representa volume da coluna esquerda no máximo e a da direita no mínimo e 100 o inverso (volumes em separado da coluna esquerda e da direita também visualizados).

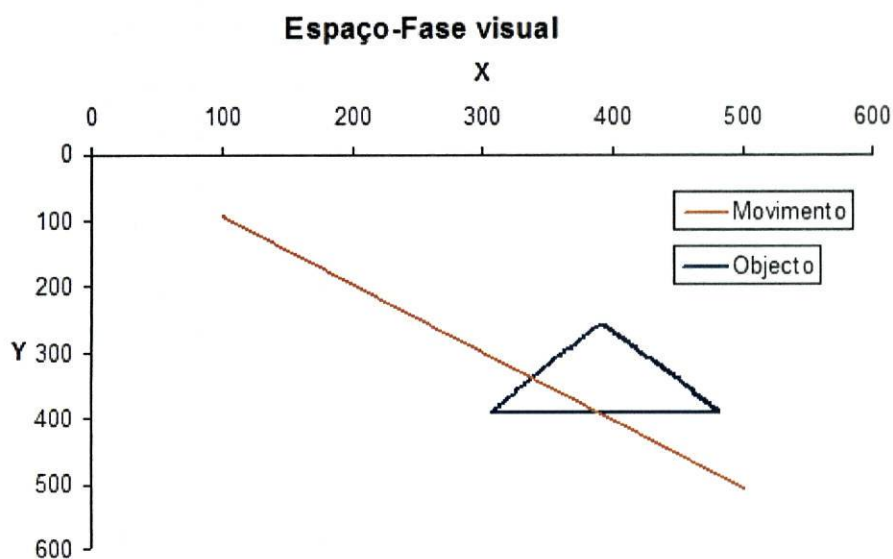


Figura 3.10: Representação das arestas do triângulo e da posição do centro-de-massa do mesmo como movimento visual num estado espacial (x, y) .

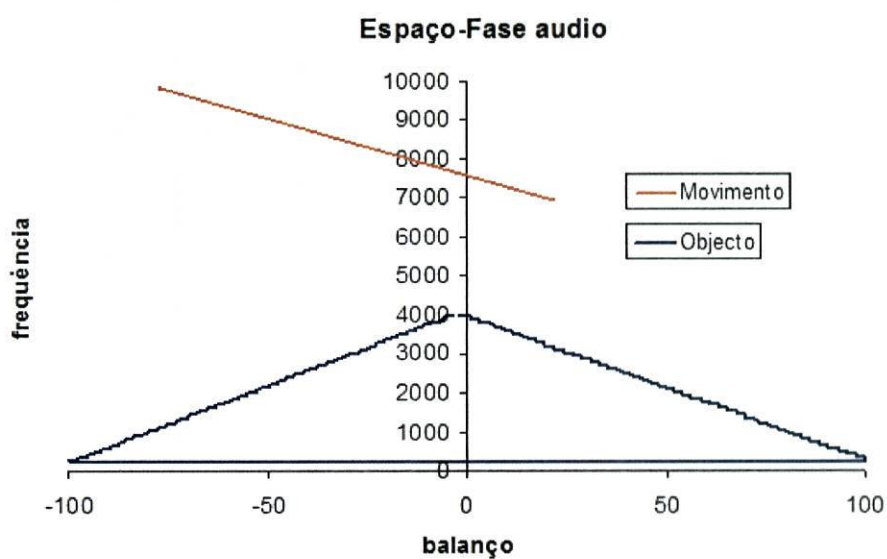


Figura 3.11: Representação das arestas do triângulo e da posição do centro-de-massa do mesmo como movimento num estado espacial do som $(frequência, balanço)$.

Capítulo 4

Detalhes de Implementação

Para realizar o modelo de sonificação proposto, implementamos uma aplicação protótipo (desenvolvido na linguagem de programação PYTHON utilizando o módulo TKINTER para codificação do GUI e do módulo TKSNAACK¹) para síntese dos canais de audio, que permite configurar o algoritmo de sonificação numa variedade de aspectos. A motivação principal para a implementação desta aplicação (MOVIE2SOUND), foi a de poder avaliar a usabilidade de algoritmos de sonificação concretos, e por forma a sugerir aplicações práticas concretas a desenvolver futuramente. Na secção 4.1 descrevemos as funcionalidades suportadas pela aplicação. Na secção 4.2 descrevemos o método de obtenção das fronteiras dos objectos. Na secção 4.3 descrevemos a conversão da forma do objecto e da sua trajectória para audio e finalmente na secção 4.4 descrevemos os diversos tipos de movimento que se pode configurar na aplicação.

4.1 Enumeração das funcionalidades da aplicação

A figura 4.1 mostra um *screenshot* da aplicação. A janela da aplicação é composta por dois painéis, o da esquerda tem a imagem a ser sonificada e o da direita mostra os *widgets* de configuração dos parâmetros da aplicação. Globalmente, o painel de configuração, os menus da GUI e as operações da aplicação, suportam e dão acesso às seguintes funcionalidades:

¹www.speech.kth.se/snack/.

- Abertura de vários tipos de ficheiros de imagem.
- Identificação das fronteiras dos objectos que compõem a imagem.
- Animação do cursor nas fronteiras dos objectos enquanto a sonificação está a ser reproduzida.
- Selecção da gama de frequências a ser utilizada para a sonificação da forma e a sonificação do movimento do centro-de-massa, incluindo a capacidade de especificar se a forma usa uma gama de frequências mais baixa ou mais alta do que a gama de frequências usada para o movimento do CM.
- Selecção se é a informação da forma, a informação do movimento do CM, ou ambos os tipos de informação que deverão ser sonificados.
- Selecção do tempo de duração da sonificação das fronteiras dos objectos.
- Selecção do número de repetições usadas na sonificação dos objectos.
- Selecção do tipo de movimento físico a realizar por um objecto, dado um conjunto de movimentos pré-definidos.
- Selecção dos parâmetros para o movimento dos objectos (e.g. posição inicial, velocidade e aceleração).

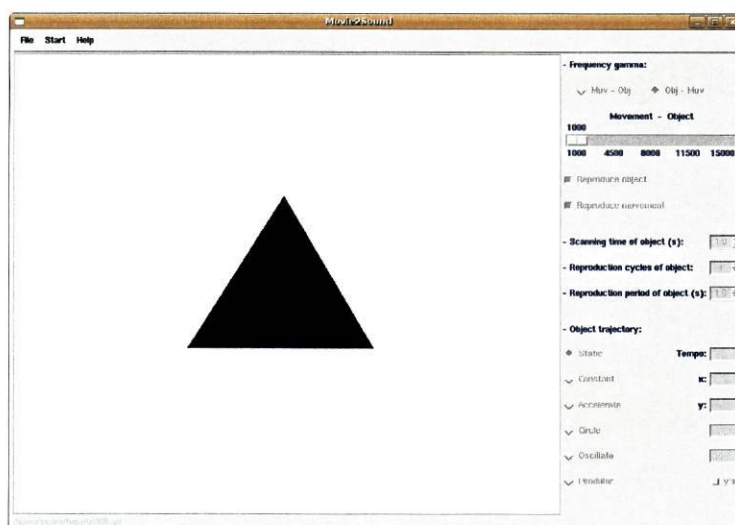


Figura 4.1: Aplicação para sonificação de imagens e animações.

Resumidamente, o software MOVIE2SOUND implementa um processo constituído por duas fases distintas mas encadeadas. Na primeira fase, é realizado a análise das fronteiras dos objectos presentes numa imagem (revelando assim de forma mais fácil a estrutura geométrica dos objectos), usando um pequeno conjunto de técnicas de processamento de imagem computacional elementares. Na segunda fase é realizada a tradução para uma onda sonora das formas dos objectos (tal como identificado na primeira fase) e também do seu movimento, caso se trate de imagens animadas (ver figura 4.2).



Figura 4.2: Diagrama de blocos da aplicação Movie2Sound.

4.2 Análise e desenho das fronteiras do objecto

A primeira fase do processamento de imagens envolve três subprocessos distintos, esquematizados na figura 4.3. Descreveremos em detalhe nas próximas secções estes subprocessos.

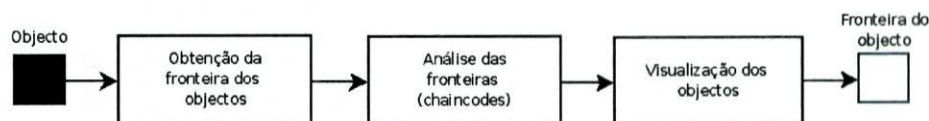


Figura 4.3: Primeira fase do Movie2Sound.

4.2.1 Obtenção da fronteira dos objectos

Este processo consiste na execução de um simples algoritmo de detecção de contraste numa imagem. Basicamente, este algoritmo baseia-se na computação da primeira derivada em relação ao movimento do espaço de uma função representando a intensidade dos pixels presentes na imagem [KSFL87]. Em particular têm-se que a primeira derivada numa região de intensidade constante é zero e assume um valor diferente de zero quando existe uma alteração do valor de intensidade dos pixels.

Matematicamente, o operador que define a magnitude e a direcção da variação da intensidade é o gradiente. Formalmente dada uma imagem representada por uma função de intensidade luminosa $f(x, y)$, o gradiente de uma imagem é definido pelo vector bidimensional apresentando em baixo:

$$\mathbf{G}[f(x, y)] = \begin{bmatrix} G_x \\ G_y \end{bmatrix} = \begin{bmatrix} \frac{\delta f}{\delta x} \\ \frac{\delta f}{\delta y} \end{bmatrix}$$

É bem conhecido da análise vectorial que o vector \mathbf{G} aponta na direcção da máxima variação de f no ponto (x, y) . (Também se usa frequentemente o símbolo ∇ para representar o símbolo do gradiente.) Na aplicação concreta de detecção de fronteiras de objectos o que mais interessa considerar é a magnitude/norma deste vector, sendo as fronteiras dos objectos associados a pontos de alto contraste. Em baixo, apresenta-se a definição formal do gradiente:

$$\|\mathbf{G}[f(x, y)]\| = [G_x^2 + G_y^2]^{1/2} = \left[\left(\frac{\delta f}{\delta x} \right)^2 + \left(\frac{\delta f}{\delta y} \right)^2 \right]^{1/2}$$

Em implementações práticas é comum substituir a norma euclideana pela norma de Manhattan como mostrado em baixo:

$$\|\mathbf{G}[f(x, y)]\| \approx |G_x| + |G_y|$$

Para aproximar as derivadas parciais $\frac{\delta f}{\delta x}$ e $\frac{\delta f}{\delta y}$ interpreta-se a função f como um mapa discreto. Assim, o valor das derivadas parciais é substituído por valores de diferença entre pixels adjacentes tal como apresentado em baixo:

$$G_x = \frac{\delta f}{\delta x} = f(x + 1, y) - f(x, y)$$

e

$$G_y = \frac{\delta f}{\delta y} = f(x, y - 1) - f(x, y)$$

Finalmente, para se obterem as fronteiras dos objectos a partir dos valores do módulo do gradiente é criada uma imagem binária, $g(x, y)$, parametrizada por um parâmetro θ definida por:

$$g(x, y) = \begin{cases} 1 & \text{se } G[f(x, y)] < \theta \\ 0 & \text{se } G[f(x, y)] \geq \theta \end{cases}$$

onde $g(x, y) = 1$ representa a cor branca e $g(x, y) = 0$ representa o preto, gerando assim uma imagem de fundo branco com as fronteiras a preto.

4.2.2 Análise das fronteiras

Neste processo são analisadas as fronteiras dos objectos existentes na imagem para obter os respectivos **Chain Codes** (códigos ou caminhos das fronteiras) [KSFL87]. Os Chain Codes são utilizados para representar a fronteira como uma sequência de segmentos de recta com uma certa direcção e dimensão. Tipicamente, esta representação é estabelecida centrando uma grelha com 4 ou 8 pontos de conectividade, como podemos visualizar na figura 4.4. O comprimento de cada segmento é definido pela resolução da grelha e as direcções são dadas pelo código escolhido. O caso mais simples, é quando a grelha de pontos de conectividade representam os pixels vizinhos do ponto central. Essa é a solução usada pela nossa implementação.

No protótipo MOVIE2SOUND, foi utilizado a grelha com 8 pontos de direcção (ver figura 4.4 (b)), para determinar o chain code dos objectos presentes na imagem. Para este efeito foi utilizado o seguinte método:

Começa-se por fazer um varrimento horizontal da imagem linha a linha. Quando se encontra um ponto pertencente a uma fronteira (pixel a preto), verifica-se se algum objecto já foi encontrado. Se não foram encontrados

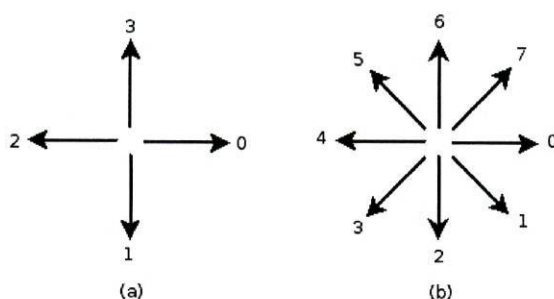


Figura 4.4: a) Chain code baseado em 4 direcções. b) Chain code baseado em 8 direcções.

objectos, é iniciado um algoritmo para obter o chain code para o objecto contendo esse ponto. Se já tiverem sido encontrados objectos, verifica-se primeiro se o ponto em análise pertence a algum dos objectos previamente encontrados. Caso isso aconteça, o ponto em análise é descartado uma vez que se considera que já foi previamente tratado. Se o ponto não pertencer a nenhum dos objectos, é iniciado um novo processo de obtenção do chain code para o objecto a que o ponto pertence (tal como acontece para o caso em que não foi encontrado qualquer objecto).

O processo de obtenção do chain code é desencadeado sempre que no varrimento da imagem linha a linha é encontrado um novo ponto pertencente a uma fronteira.

Este processo para obtenção do chain code, começa por guardar um conjunto de informação, designadamente:

- As coordenadas do ponto inicial, x_i^k e y_i^k ;
- As coordenadas mínimas e máximas do objecto, x_{min}^k , x_{max}^k , y_{min}^k e y_{max}^k (no início são iguais ao x_i^k e y_i^k).

k - é um índice para o objecto a ser identificado (1, 2, ..., n).

De seguida, a partir do ponto em análise (x_i^k, y_i^k) são observados os pontos vizinhos, distinguindo-se dois casos (ver figura 4.5):

- Observam-se os pontos vizinhos no sentido horário (0, 1, 2, 3, 4, 5, 6, 7), parando-

se no primeiro vizinho com pixel a preto, que se assume pertencer à fronteira do objecto. Este sentido de observação é usado para descrever fronteiras com declives negativos, usando o menor número possível de pixels. No caso em que o vizinho de fronteira encontrado se encontre nas posições (5, 6, 7), a próxima pesquisa de vizinhos será realizada no sentido contrário (caso em baixo).

- Observam-se os pontos vizinhos no sentido anti-horário (0, 7, 6, 5, 4, 3, 2, 1), parando-se no primeiro vizinho com pixel a preto, que igualmente se assume pertencer à fronteira do objecto. Este sentido de observação é usado para obter descrições de fronteiras com declives positivos. Caso o vizinho da fronteira encontrado esteja nas posições (1, 2, 3). a próxima pesquisa de vizinhos é realizada no sentido horário (caso em cima).

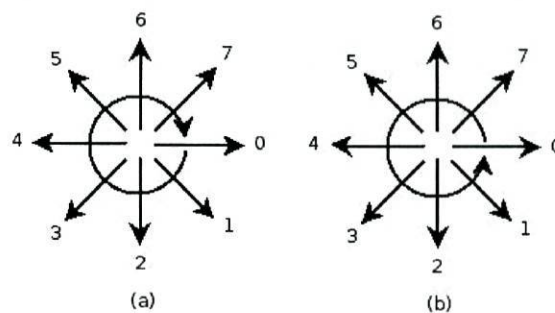


Figura 4.5: Sentidos de observação dos pixels vizinhos: a) Sentido horário. b) Sentido anti-horário.

A figura 4.6 ilustra o processo de construção de chain codes. Esta solução heurística permite que arestas/fronteiras de objectos com o mesmo comprimento geométrico gerem aproximadamente o mesmo número de pontos no chain code, independentemente da sua direcção.

Este processo termina quando o ponto em análise for novamente o ponto inicial, x_i^k e y_i^k , identificando-se assim um ciclo no processo de construção do chain code.

Após a obtenção do chain code do objecto, é calculado e memorizado o ponto que representa o centro-de-massa do objecto, de acordo com as seguintes expressões:

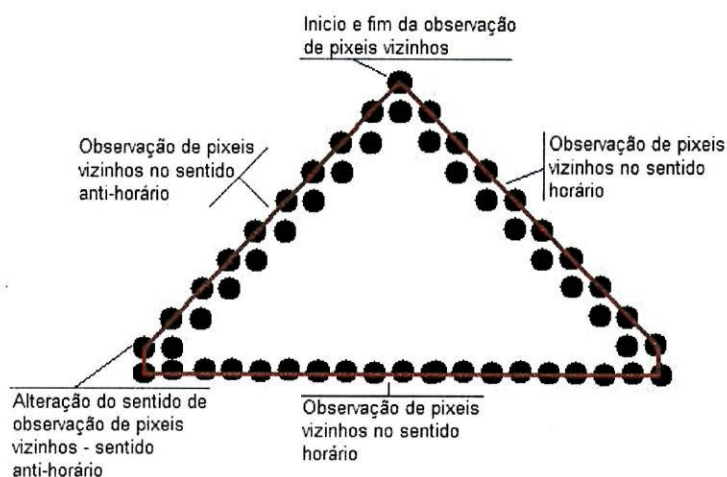


Figura 4.6: Descrição do processo de construção de chain codes.

$$\left\{ \begin{array}{l} x_{cm} = \frac{\sum_{i=0}^n x_i}{n} \\ y_{cm} = \frac{\sum_{i=0}^n y_i}{n} \end{array} \right.$$

n representa o número de pontos mapeados.

4.2.3 Visualização de objectos

Para facilitar o *debug* do sistema, é visualizado informação que indica qual o conjunto de pontos pertencente à fronteira de um certo objecto. Em particular, todos os pontos de um certo objecto são visualizados na GUI com uma determinada cor. (Isto é implementado associando uma única chave (*tag*) do módulo TKINTER a todos esses pontos).

4.3 Tradução da forma do objecto e do seu movimento para som

A segunda fase do processo de sonificação é constituída por quatro processos, esquematizados na figura 4.7.

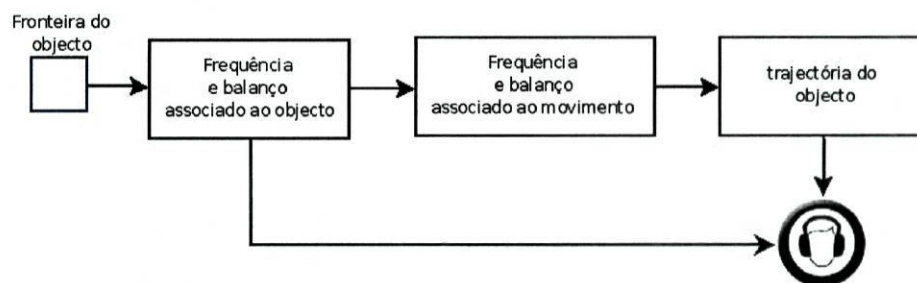


Figura 4.7: Segunda fase do Movie2Sound.

4.3.1 Cálculo de frequências e níveis de balanço para sonificação de objecto

Este processo inicial começa por calcular um conjunto de parâmetros utilizados no processo de sonificação, levando em conta a gama de frequências parametrizada para a reprodução acústica do objecto $[f_{min_obj}, f_{max_obj}]$. Note-se que a gama de frequências utilizada para sonificar formas de objectos poderá ser diferente da gama de frequências utilizada para a sonificação do movimento. Em particular, determina a constante kv_{obj} que indica a variação em frequência por unidade de deslocamento no eixo dos y 's nos cursores de sonificação, isto é, se um cursor se desloca um pixel para cima, a frequência da onda audio produzida é incrementada de kv_{obj} e caso o cursor se desloca um pixel para baixo, a frequência é decrementada de kv_{obj} . kv_{obj} é dado pela seguinte expressão:

$$kv_{obj} = \frac{f_{max_obj} - f_{min_obj}}{y_{max} - y_{min}}$$

Em cima, y_{max} e y_{min} representam as coordenadas máximas e mínimas no eixo dos y 's para cada objecto (tal como calculado pelo algoritmo de deteção dos chain codes).

Neste processo inicial é também calculado uma constante kh_{obj} que representa a variação do nível do balanço usado na produção das ondas sonoras, nomeadamente deslocamentos horizontais de um pixel para a direita fazem aumentar o nível da coluna/auricular direita de um valor kh_{obj} e diminuir o mesmo valor de kh_{obj} na coluna/auricular da esquerda. Inversamente, movimentos dos cursores de sonificação para a esquerda de um pixel fazem aumentar o nível da coluna/auricular da esquerda de kh_{obj} e diminuir o mesmo valor de kh_{obj} na coluna/auricular da direita. kh_{obj} é dado pela seguinte expressão:

$$kh_{obj} = \frac{50.0}{\frac{x_{max} - x_{min}}{2}}$$

Em cima, x_{max} e x_{min} representam as coordenadas máximas e mínimas no eixo dos x 's para cada objecto (tal como calculado pelo algoritmo de deteção dos chain codes).

A figura 4.8, ilustra as posições máximas e mínimas necessárias para o cálculo das constantes kh_{obj} e kv_{obj} .

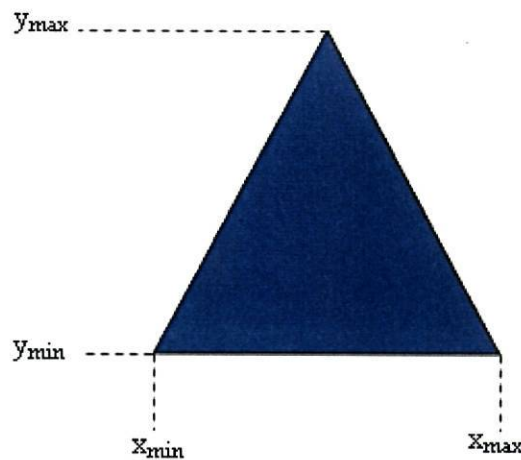


Figura 4.8: Cálculo das constantes kh_{obj} e kv_{obj} .

A título de exemplo, considere que o volume é de 50 % do valor máximo nas duas colunas quando a abcissa do ponto P_i é igual à abcissa do ponto intermédio entre os extremos do objecto (x_{max} e x_{min}) e vai aumentando até 100 % numa coluna e diminuindo até 0 % na outra coluna quando a abcissa do ponto P_i é igual à abcissa máxima ou mínima (x_{max} ou x_{min}). Neste caso, deslocamentos do cursor de fronteira

para a direita faria aumentar o volume da coluna do lado direito para um valor entre 50 % e 100 %. Note-se que a soma da percentagem do volume das duas colunas em qualquer ponto é sempre igual a 100 %.

Depois de calculadas as constantes kh_{obj} e kv_{obj} , são ainda pré-gerados três vectores contendo a informação da onda sonora a gerar em cada instante: um com as frequências em cada ponto; um com a percentagem do volume para a coluna esquerda e por último, um com a percentagem do volume para a coluna direita.

As frequências do som emitido em cada ponto são obtidas da seguinte forma:

Ao primeiro ponto da fronteira mapeado é atribuída a frequência máxima do objecto (f_{max_obj}), pois é o ponto de ordenada máxima (y_{max}). Depois verifica se a direcção, evidenciada no chain code, associada ao ponto seguinte for igual a 1, 2 ou 3, então a frequência desse ponto é decrementada de kv_{obj} Mhz. Se a direcção, evidenciada no chain code, associada ao ponto seguinte for igual a 5, 6 ou 7, então a frequência desse ponto é incrementada de kv_{obj} Mhz. Se a direcção, evidenciada no chain code, associada ao ponto seguinte for igual a 0 ou 4, então a frequência desse ponto é igual à frequência do ponto anterior.

Este processo repete-se até se ter percorrido todo o chain code do objecto, obtendo-se assim um vector com as frequências do som atribuídas a cada ponto da fronteira do objecto.

O gráfico da figura 4.9, apresenta um exemplo de como os valores da frequência de som variam quando a posição do cursor se desloca entre y_{min} e y_{max} para um qualquer objecto.

As percentagens dos volumes em cada ponto são obtidas da seguinte forma:

Para cada ponto pertencente à fronteira do objecto, verifica se a abcissa do ponto P_i se encontra à direita ou à esquerda da abcissa do ponto de centro de massa (x_{cm}).

$$V_{lo} = \begin{cases} 50.0 & \text{se } x_i = x_{cm} \\ 50.0 + (x_{cm} - x_i) * kh_o & \text{se } x_i \neq x_{cm} \end{cases}$$

$i = 0, \dots, n.$

e

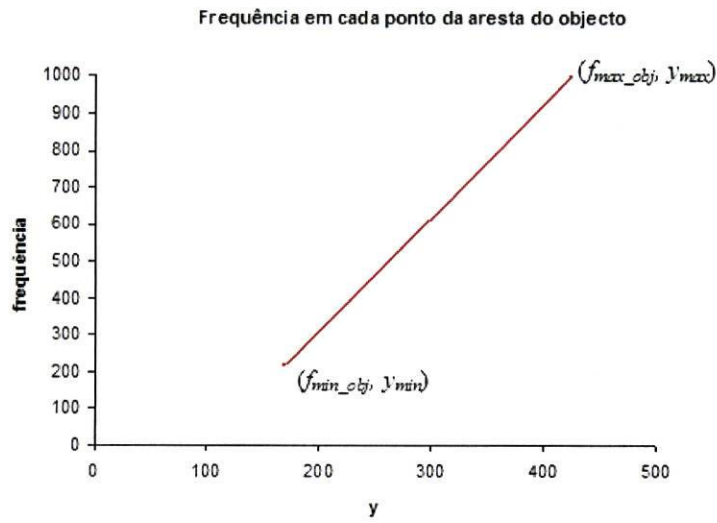


Figura 4.9: Frequências entre os pontos y_{\min} e y_{\max} para um objecto.

$$V_{ro} = \begin{cases} 50.0 & \text{se } x_i = x_{cm} \\ 50.0 + (x_i - x_{cm}) * kh_o & \text{se } x_i \neq x_{cm} \end{cases}$$

$i = 0, \dots, n.$

A figura 4.10, ilustra o método acima descrito.

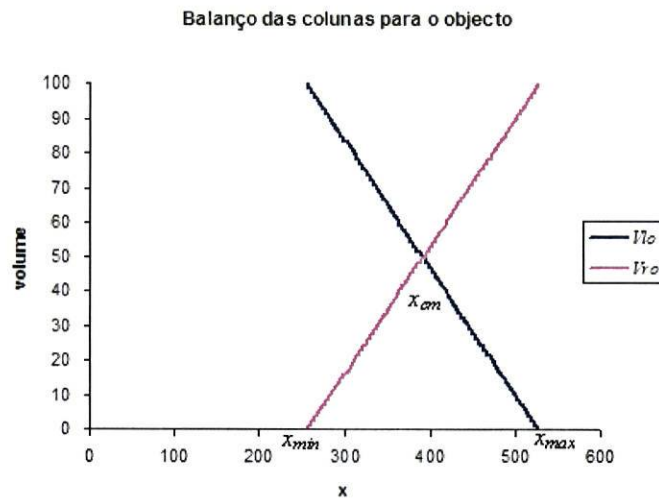


Figura 4.10: Balanço do volume entre os pontos x_{\min} e x_{\max} para um objecto.

Este processo termina quando todos os pontos da fronteira tenham sido percorridos, obtendo assim dois vectores com as percentagens dos volumes associados às respectivas colunas (V_{lo} - coluna da esquerda; V_{ro} - coluna da direita), atribuídas por cada ponto da fronteira do objecto.

4.3.2 Cálculo de frequências e níveis de balanço para sonificação do movimento

Este processo começa por calcular um conjunto de parâmetros utilizados no processo de sonificação, levando em conta a gama de frequências parametrizada para a reprodução acústica do movimento $[f_{min_mov}, f_{max_mov}]$. Note-se que a gama de frequências utilizada para sonificar o movimento poderá ser diferente da gama de frequências utilizada para a sonificação das formas dos objectos. Em particular, determina a constante kv_{mov} que indica a variação em frequência por unidade de deslocamento no eixo dos y 's da posição do centro-de-massa do objecto de sonificação, isto é, se a posição do centro-de-massa do objecto se desloca um pixel para cima, a frequência da onda audio produzida é incrementada de kv_{mov} e caso a posição do centro-de-massa do objecto se desloca um pixel para baixo, a frequência é decrementada de kv_{mov} . kv_{mov} é dado pela seguinte expressão:

$$kv_{mov} = \frac{f_{max_mov} - f_{min_mov}}{H}$$

Em cima H representa a altura em pixels da área onde o objecto se vai deslocar².

Neste processo é também calculado uma constante kh_{mov} que representa a variação do nível do balanço usado na produção das ondas sonoras. kh_{mov} é dado pela seguinte expressão:

$$kh_{mov} = \frac{50\%}{W}$$

Em cima W representa metade da largura em pixels da área onde o objecto se vai deslocar³.

²Na instalação onde desenvolvemos e testamos a aplicação, usamos H igual a 664.

³Na instalação onde desenvolvemos e testamos a aplicação, usamos W igual a 400.

A título de exemplo, considere que o volume é de 50 % nas duas colunas quando a abcissa do ponto de centro de massa do objecto é igual à abcissa 400 (ponto intermédio entre os extremos do plano onde o objecto se vai mover) e vai aumentando até 100 % numa coluna e diminuindo até 0 % na outra coluna quando a abcissa de centro de massa do objecto é igual às abcissas 0 ou 800 (abcissa mínima ou máxima do plano onde o objecto se vai mover).

À semelhança do que acontece na sonificação da forma, são pré-gerados três vectores contendo a informação da onda sonora a gerar em cada instante: um com as frequências em cada ponto da trajectória; um com a percentagem do volume para a coluna esquerda e por último, um com a percentagem do volume para a coluna direita.

As frequências do som emitido em cada ponto da trajectória são obtidas da seguinte forma:

Ao ponto inicial da trajectória é atribuída a frequência segundo a expressão:

$$f_{mov_i} = f_{min_{mov}} + \frac{f_{max_{mov}} - f_{min_{mov}}}{2} - \left(y_i - \frac{H}{2}\right) * kv_m$$

$i = 0$

Durante um movimento de um objecto, o valor de frequência e balanço da onda sonora produzida é constantemente actualizado para fazer reflectir as sucessivas posições do centro-de-massa. Considerando que o objecto se desloca na vertical no valor de Δy pixels, faz-se a actualização da frequência segundo a equação apresentada em baixo:

$$f_{mov_t} = f_{mov_{t-1}} - \Delta y * kv_{mov}$$

$t = 1, \dots, n$

A figura 4.11, ilustra o método acima.

As percentagens dos volumes/balanço em cada ponto da trajectória são obtidas levando em conta a posição horizontal do centro-de-massa x_i , segundo a equação em baixo:

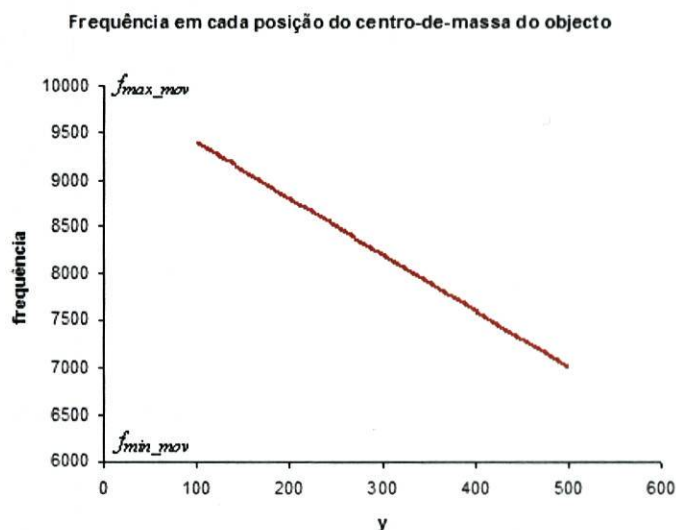


Figura 4.11: Frequências entre os pontos inicial e final de uma trajetória.

$$V_{lm} = \begin{cases} 50.0 & \text{se } x_t = W \\ 50.0 + (W - x_t) * kh_{mov} & \text{se } x_t \neq W \end{cases}$$

$t = 0, \dots, n.$

e

$$V_{rm} = \begin{cases} 50.0 & \text{se } x_t = W \\ 50.0 + (x_t - W) * kh_{mov} & \text{se } x_t \neq W \end{cases}$$

$t = 0, \dots, n.$

A figura 4.12, ilustra o método acima descrito.

4.3.3 Reprodução acústica do objecto e da sua trajetória

É também possível com o software MOVIE2SOUND seleccionar se se pretende reproduzir acusticamente apenas a forma do objecto, apenas a trajetória ou ambos, designadamente, primeiro a forma e depois a trajetória do objecto.

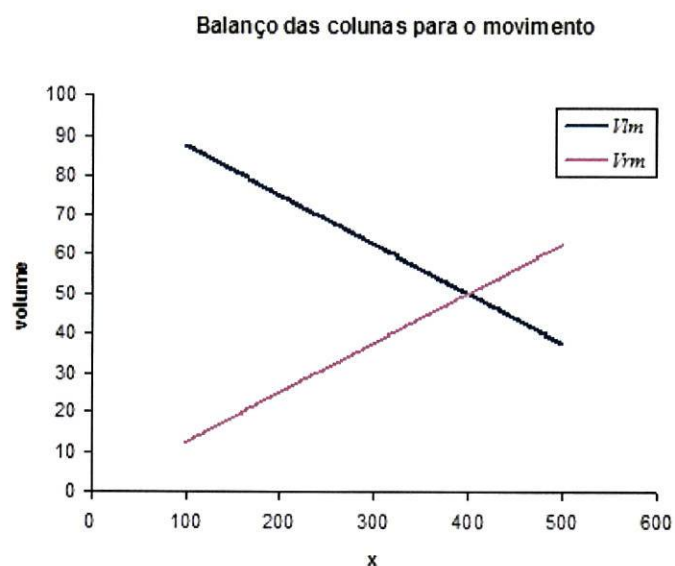


Figura 4.12: Balanço do volume entre os pontos inicial e final de uma trajetória.

Quando se selecciona a reprodução apenas da forma do objecto, pode-se parametrizar o tempo de reprodução do objecto, ou seja, o tempo de varrimento da forma (fronteira) do objecto, e o número de vezes que a forma do objecto é reproduzida. No intermeio entre dois varrimentos o algoritmo de sonificação faz sempre uma pausa de dois segundos.

Quando se escolhe a reprodução apenas da trajetória do objecto, pode-se parametrizar o tipo de movimento que pretende visualizar (estático, constante, acelerado, circular, oscilante e pendular) com os respectivos dados parametrizáveis associados a cada tipo de movimento (descrito na secção 4.4).

Quando se selecciona a reprodução combinada da forma e trajetória do objecto, pode-se parametrizar os dois conjuntos de parâmetros associados. Nestes casos, a forma do objecto é sonificada em primeiro lugar seguido da reprodução de um marcador sonoro (um beep) que serve para indicar que a sonificação da forma do objecto foi concluída. Posteriormente faz-se a sonificação da trajetória do centro-de-massa do objecto.

4.4 Trajectórias pré-definidas para os objectos

Nesta secção apresentamos as equações de movimento, tal como usamos no teste da aplicação. Usamos apenas movimentos simples tais como estudados em livros de texto de introdução à física/mecânica. Designadamente implementamos os seguintes tipos de movimento: movimento linear uniforme, movimento uniformemente acelerado, movimento circular, movimento oscilante nas direcções horizontal e vertical, movimento pendular, bem como posições estáticas.

Nas descrições apresentadas em baixo, Δx e Δy , representam o deslocamento no eixo dos x 's e dos y 's respectivamente, para cada instante t .

4.4.1 Posição estática

Quando os objectos estão estáticos, a única transformação envolvida é o deslocamento do objecto para uma certa posição (x, y) (ver figura 4.13). Neste caso o valor de Δx e Δy leva em conta apenas a posição desejada (x, y) e a posição original do centro-de-massa (x_{cm}, y_{cm}) no instante $t = 0$ no referencial fixo no centro do ecrã. Apresenta-se em baixo a expressão para o cálculo do Δx e Δy :

$$\Delta x = x - x_{cm} \text{ e } \Delta y = y - y_{cm}$$



Figura 4.13: Posição estática.

4.4.2 Movimento rectilíneo uniforme

Este tipo de movimento corresponde ao movimento rectilíneo uniforme (ver figura 4.14). Neste caso, é possível parametrizar a posição inicial do objecto (x_0, y_0) , a velocidade v_x e v_y e a duração durante a qual o objecto é animado T .

Inicialmente, o objecto é deslocado para uma posição inicial (x_0, y_0) e de seguida é calculado os valores de Δx e Δy da seguinte forma:

$$\begin{cases} \Delta x = v_x * t - v_x * (t - 1) = v_x \\ \Delta y = v_y * t - v_y * (t - 1) = v_y \end{cases} \\ t = 1, \dots, T$$

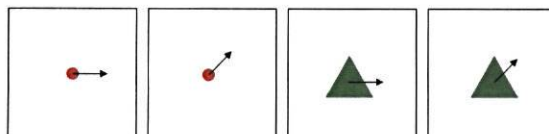


Figura 4.14: Movimento rectilíneo uniforme.

4.4.3 Movimento rectilíneo uniformemente acelerado

Este tipo de movimento corresponde ao movimento rectilíneo uniformemente acelerado (ver figura 4.15). Neste caso, é possível parametrizar a posição inicial do objecto (x_0, y_0) , a aceleração a_x e a_y e a duração durante a qual o objecto é animado T .

Inicialmente, o objecto é deslocado para uma posição inicial (x_0, y_0) e de seguida é calculado o Δx e Δy da seguinte forma:

$$\begin{cases} \Delta x = (1/2) * a_x * t^2 - (1/2) * a_x * (t - 1)^2 \\ \Delta y = (1/2) * a_y * t^2 - (1/2) * a_y * (t - 1)^2 \end{cases} \\ t = 1, \dots, T$$

4.4.4 Movimento circular

Este tipo de movimento corresponde ao movimento circular uniforme (ver figura 4.16). Neste caso, é possível parametrizar a posição inicial do objecto (x_0, y_0) , o período do movimento T , o raio r e a duração T_p .

Inicialmente, o objecto é deslocado para uma posição inicial (x_0, y_0) e de seguida é calculado o Δx e Δy da seguinte forma:

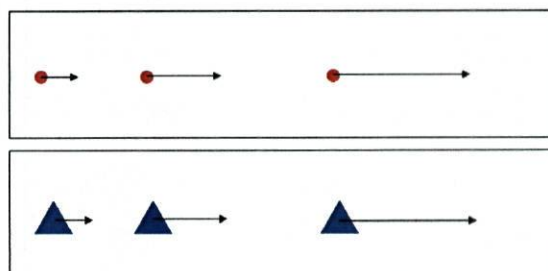


Figura 4.15: Movimento rectilíneo uniformemente acelerado.

$$\begin{cases} \Delta x = r * \cos(\omega * t) - r * \cos(\omega * (t - 1)) \\ \Delta y = r * \sin(\omega * t) - r * \sin(\omega * (t - 1)) \end{cases}$$

Sendo $t = 1, \dots, T_p$ e $\omega = \frac{2\pi}{T}$.

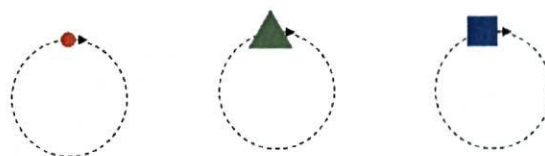


Figura 4.16: Movimento circular.

4.4.5 Movimento oscilante

Este tipo de movimento corresponde ao movimento harmónico simples (movimento de mola elástica) no eixo dos x 's ou no eixo dos y 's (ver figura 4.17). Neste caso, é possível parametrizar a posição inicial do objecto (x_0, y_0) , o deslocamento inicial num dos eixos Δd , a constante elástica da mola k e o tempo T . Despezamos o efeito da inércia no movimento (alto valor de força de atrito).

Inicialmente, o objecto é colocado na posição inicial $(x_0 + \Delta d, y_0)$, para oscilações horizontais ou $(x_0, y_0 + \Delta d)$ para oscilações verticais. Em cada instante, é calculado o valor de Δx , ou Δy (conforme o tipo de movimento seleccionado), calculando a aceleração do objecto, definida como $a = -k * \Delta d$. Os deslocamentos em cada instante são dados pela expressão:

$$\Delta x_t = \Delta d = a$$

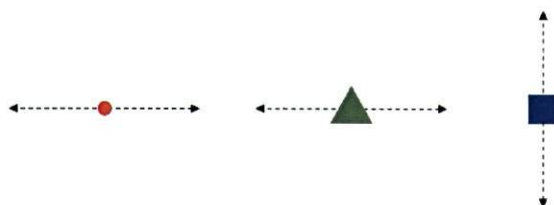


Figura 4.17: Movimento oscilante.

para oscilações horizontais, e

$$\Delta y_t = \Delta d = a$$

para oscilações verticais.

4.4.6 Movimento pendular

Este tipo de movimento corresponde ao movimento pêndular sem rotação completa (ver figura 4.18). Neste caso, pode-se parametrizar a posição de fixação do pêndulo (x_0, y_0) , o ângulo máximo de abertura θ_{max} , que é também o ângulo inicial, o comprimento do pêndulo l e a duração da animação T .

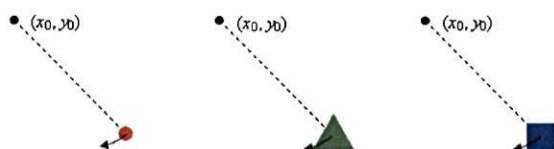


Figura 4.18: Movimento pendular.

Inicialmente, o objecto é colocado na posição $(x_0 + l * \cos(\theta_{max}), y_0 + l * \sin(\theta_{max}))$. Δx e Δy são actualizados calculando a aceleração angular, $a = -\sin(\theta_t)$. De seguida calcula-se o ângulo θ_t que o objecto vai fazer em cada instante. Ignora-se o momento inercial do pêndulo.

$$\theta_t = \begin{cases} \theta_{max} & \text{para } t = 0 \\ \theta_{anterior} + a * t & \text{para } t > 0 \end{cases}$$

Os deslocamentos do pêndulo a cada instante é dado por:

$$\begin{cases} \Delta x = l * \cos(\theta_t) - l * \cos(\theta_{t-1}) \\ \Delta y = l * \sin(\theta_t) - l * \sin(\theta_{t-1}) \end{cases}$$

Capítulo 5

Avaliação da Usabilidade do Software

Para avaliar a usabilidade da solução proposta, realizamos dois protocolos experimentais. Foi realizado um primeiro protocolo experimental piloto com um único sujeito com deficiências visuais. Este primeiro teste teve o intuito de fazer uma avaliação provisória da viabilidade da solução e ter uma compreensão empírica de qual o tipo de imagens e estímulos seria de mais fácil identificação. O segundo protocolo experimental envolveu três sujeitos cegos e amblíopes e teve como objectivo principal avaliar a robustez dos dados obtidos no primeiro teste e averiguar do nível de heterogeneidade na resposta de diferentes sujeitos.

Os sujeitos foram testados com uma variedade de estímulos visuais e a sua respectiva sonificação usando a aplicação desenvolvida. Utilizamos um desenho experimental multi-factor, experimentando com imagens de várias formas geométricas elementares (primeiro factor) e fazendo as formas moverem-se no plano segundo algum tipo de movimento elementar (segundo factor). As formas geométricas utilizadas foram: um ponto (para identificação do centro-de-massa de um objecto), figura quadrangular, triângulo, círculo/oval e um hexágono. Os tipos de movimento utilizados foram: estático (sem movimento), movimento com velocidade uniforme (horizontal, vertical ou oblíqua com alguma inclinação), movimento com aceleração uniforme, movimento circular uniforme, oscilações horizontais ou verticais (tipo uma mola elástica) e movimento pendular. Todas as cenas sonificadas continham apenas um objecto.

Para cada estímulo apresentado, solicitamos ao sujeito/utilizador que desenhasse em papel a forma do objecto sonificado e indicasse o tipo de movimento realizado, baseando-se apenas na informação presentes nos sons reproduzidos. Para os dois tipos de resposta pretendidos (forma e movimento), solicitamos também aos sujeitos que nos indicasse um nível de confiança numa escala de +3 — máxima confiança, a -3 máxima incerteza. Os estímulos foram reproduzidos um certo número de vezes sendo repetido sempre e caso o sujeito o solicitasse.

Testamos os sujeitos em três cenários:

- Apenas as arestas dos objectos foram sonificadas.
- Apenas a posição e o movimento do centro-de-massa do objecto foi sonificado.
- Sonificação de ambos os tipos de informação, forma e movimento do centro-de-massa do objecto (apresentando primeiro a informação audio da forma e depois da posição/movimento).

Na secção 5.1 apresentamos os resultados do primeiro protocolo experimental piloto efectuado a um único sujeito e na secção 5.2 apresentamos os resultados do segundo protocolo experimental efectuado a três sujeitos.

5.1 Primeiro protocolo experimental

Seguindo o desenho experimental descrito na introdução, testamos o sujeito numa sequência de estímulos de acordo com que nos pareceu ser um incremento progressivo do nível de dificuldade. Testamos primeiro apenas as formas dos objectos, de formas mais simples para formas mais complicadas (da esquerda para a direita na tabela 5.1), de seguida testamos o movimento do centro-de-massa dos objectos, indo de movimentos mais simples para movimentos mais complicados (de cima para baixo na tabela 5.1), e finalmente num sub-conjunto de combinações de forma \times movimento. A tabela 5.1, apresenta os resultados que iremos descrever seguidamente. As células da tabela 5.1 com fundo mais escuro indicam respostas consideradas incorrectas.

Os resultados mostram que o utilizador foi capaz de discriminar as formas a partir do som reproduzido para três das formas geométricas utilizadas (rectângulo, triângulo e

circulo). O utilizador não conseguiu inferir correctamente a forma mais complicada utilizada (hexágono). Isto poderá ser explicado devido à dificuldade de identificar as arestas oblíquas do hexágono, que dadas as configurações da aplicação usadas, fez com que a variação na escala do balanço e frequência não fosse suficientemente grande para que as arestas oblíquas fossem correctamente apreendidas. Os resultados mostram também que a confiança em todas as respostas correctas foi alta, mas diminuiu com o aumento da complexidade da forma do objecto. registou-se também que a confiança atribuída pelo sujeito diminuiu com o aumento da suposta complexidade do objecto, tendo sido +3 para o rectângulo, +2 para o triângulo e +1 para a figura oval.

O movimento pontual do centro-de-massa foi também correctamente percebido na maioria dos casos — horizontal, vertical e oblíquo. Foi registado um aumento da confiança de +1 para +2, nas primeiras duas tentativas, e para +3 nas tentativas restantes, sugerindo assim algum nível de adaptabilidade pelo utilizador devido ao ganho de familiaridade com os estímulos. Os movimentos linearmente acelerados foram correctamente percebidos na direcção mas não o efeito da aceleração. O movimento circular foi também confundido com o movimento oscilatório, o que foi inesperado uma vez que já tinha sido registado uma correcta identificação da forma oval. As oscilações verticais e horizontais e o movimento pendular foram correctamente percebidos.

Sonificações combinadas da forma e movimento foram também correctamente percebidas em todas as tentativas, com alta confiança nos dois aspectos, forma e movimento. A excepção foi apenas a primeira experiência testada (triângulo a mover-se para baixo), em que na primeira tentativa foi confundido com um circulo a mover-se para baixo. A repetição do teste resultou numa correcta resposta do utilizador.

5.2 Segundo protocolo experimental

No segundo protocolo experimental, utilizamos três sujeitos¹. Neste conjunto de testes executamos um protocolo experimental semelhante ao protocolo piloto. Os sujeitos apresentavam idades entre os 48 e os 50 anos, tendo dois deles como formação académica o 9º ano de escolaridade e o terceiro o 6º ano. Nenhum dos sujeitos é cego

¹Agradecemos à ACAPO (Associação dos Cegos e Amblíopes de Portugal), o facto de nos ter facilitado o acesso e marcado entrevista a três sujeitos cegos ou amblíopes, e disponibilizado as suas instalações para realização do protocolo experimental





















Tipo de movimento	(CM)				
Estático	↖ NW -1 ↘ SE -2 ↙ SW -3 ← W -3	 -3 5 s 3x	 -2 5 s 3x	 +1 5 s 3x	
Velocidade constante	→ $V_x=1, V_y=0$ -3 ↓ $V_x=0, V_y=1$ -3 ↘ $V_x=1, V_y=1$ -3	 → -2 / +3 5 s 1x 10 s  ↓ -3 / ? 5 s 1x 10 s -3 / +3	 ↑ +2 / +3 5 s 1x 10 s		
Acelerado					
Circular					
Oscilação vertical	 -3 10 s				
Oscilação horizontal	 -3 10 s		 +3 5 s 1x 10 s  +3		
Movimento pendular	 +2 10 s	 -3 5 s 1x 10 s  +3			

Tabela 5.1: Resultados do protocolo experimental com um único sujeito (amblíope, formação académica superior - 3º ano, 21 anos de idade). Nas sucessivas colunas da primeira linha estão representadas as diferentes formas geométricas usadas (ponto, quadrado/rectângulo, triângulo, círculo/oval, e hexágono). Nas sucessivas linhas da primeira coluna são indicados os diversos tipos de movimento (estático, velocidade uniforme, acelerado, movimento circular, oscilação vertical e horizontal, e movimento pendular). As respostas do sujeito aparecem dentro das células. A figura dentro das células indica a resposta tal como desenhada no papel ou gesticulada pelo sujeito, e as setas representam a direcção ou o tipo de movimento. As células com fundo mais escuro indicam respostas incorrectas. Valores na escala $[-3, 3]$ indica a confiança do sujeito na sua resposta (+3 máxima confiança, -3, máxima incerteza). Valores duplos de confiança separados por uma barra, representa respectivamente a confiança na forma e no movimento. No canto superior direito de cada célula contém informação especificando a duração da sonificação da forma, o número de vezes que o objecto foi sonificado e a duração da informação de movimento. As células vazias representam configurações não apresentadas. Valores de V_x e V_y indicam valores de velocidade no eixo do x e do y e os valores a_x e a_y representam acelerações (pixels/t). [Ver texto para explicação detalhada.]

de nascença. As tabelas 5.2, 5.3 e 5.4 apresentam os resultados obtidos.

Os resultados mostram que, de uma forma geral, os utilizadores foram capazes de discriminar as formas a partir do som reproduzido para três das formas geométricas utilizadas (rectângulo, triângulo e círculo), com excepção do último sujeito que apenas foi capaz de discriminar o círculo. Os utilizadores não conseguiram inferir correctamente a forma mais complicada (hexágono), excepto o primeiro sujeito mas com um nível de confiança de -2 (muita incerteza). Os resultados mostram também que a confiança em todas as respostas correctas foi alta, mas diminuiu com o aumento da complexidade da forma do objecto.

O movimento pontual do centro-de-massa foi também e de uma forma geral correctamente identificado na maioria dos casos — horizontal, vertical e oblíquo, com a única excepção do segundo sujeito (tabela 5.3) que confundiu o movimento oblíquo com o movimento curvilíneo. O movimento vertical linearmente acelerado de baixo para cima não foi correctamente percebido. O movimento horizontal linearmente acelerado foi correctamente percebido na direcção mas não o efeito da aceleração. O movimento pendular foi confundido com o movimento oscilatório por dois dos sujeitos. Um dos sujeitos confundiu o movimento pendular com o movimento curvilíneo complexo (tabela 5.2). As oscilações verticais e horizontais foram correctamente percebidas.

No que concerne às sonificações combinadas da forma e movimento, apenas o segundo sujeito conseguiu inferir quase a totalidade dos estímulos apresentados. Este sujeito errou na experiência em que o rectângulo se move pendularmente, tendo confundido o movimento pendular com o movimento horizontal oscilatório, e errou na experiência do rectângulo a mover-se para baixo, em que o movimento para baixo foi confundido com um movimento horizontal para a esquerda. O primeiro sujeito apenas conseguiu inferir correctamente a experiência do rectângulo a mover-se horizontalmente para o lado direito. Nas restantes experiências, o primeiro sujeito confundiu o triângulo com o círculo e o quadrado com o triângulo. Confundiu também o movimento oscilatório horizontal com o pendular. Este utilizador alegou uma acumulação de cansaço à medida que o protocolo experimental se aproximava do final, o que poderá explicar parte do aumento da taxa de insucessos. Com o terceiro sujeito, apenas se testou um rectângulo a mover-se para a direita e outro para baixo. Em ambas as experiências o sujeito não conseguiu inferir correctamente o movimento apresentado, confundindo o movimento para baixo com um movimento para a direita e o movimento para a

direita com um movimento para a direita mas fazendo uma pequena curva para baixo no final.














Tipo de movimento	(CM)				
Estático		 -3	 -3	 +1  -2	 -2
Velocidade constante	$\rightarrow V_x = 1, V_y = 0$ 0 $\downarrow V_x = 0, V_y = 1$ -3 $\swarrow V_x = 1, V_y = 1$ -3	 \rightarrow -3 / -3	 \rightarrow -2 / -2		
Acelerado	$\rightarrow a_x = 0,2$ -3 $\rightarrow a_x = 0,2$ +3				
Circular					
Oscilação vertical	\leftrightarrow -2 \updownarrow +3				
Oscilação horizontal	\leftrightarrow -3		 -1		
Movimento pendular		 -1			

Tabela 5.2: Resultados do primeiro sujeito (cego, formação académica - 9º ano, 48 anos de idade) do segundo protocolo experimental. Estrutura da tabela e os dados nela presente seguem as mesmas convenções que a tabela 5.1. [Ver texto para explicação detalhada.]

A partir dos tempos de resposta (não registados), constatamos algum nível de adaptabilidade do utilizador, sugerindo assim uma rápida curva de aprendizagem ou ganho de familiaridade pelos sujeitos. Como em nenhum dos casos se conseguiu transferir efectivamente informação sobre aceleração, iremos futuramente investigar formas de distinguir mais claramente os movimentos acelerados dos movimentos com velocidade uniforme. Um dos sujeitos sugeriu também duas modificações no sistema de sonificação: produção de um som discreto (*beep*) na passagem por arestas de objectos e produzir um som interrompido para indicar graus de curvatura nas linhas curvilíneas. Iremos futuramente também avaliar a praticabilidade destas duas sugestões. Pelo que foi dito acima e pelas tabelas, pode-se verificar que o sujeito do primeiro protocolo experimental demonstrou uma maior taxa de sucesso e nível de confiança das suas














Tipo de movimento	(CM)				
Estático		 +3 10 s 3x	 +3 10 s 2x	 +2 10 s 3x	 +3 10 s 3x
Velocidade constante	$\rightarrow V_x = 1, V_y = 0$ +3 2x $\downarrow V_x = 0, V_y = 1$ +3 1x $\swarrow V_x = 1, V_y = 1$ +2 3x	 +3 / +3 8 s  +3 / ? 8 s 1x 10 s	 +3 / ? 8 s 1x 10 s		
Acelerado	$? a_x = 0,2$ -3 10 s $\rightarrow a_x = 0,2$ +3 10 s				
Circular					
Oscilação vertical	?				
Oscilação horizontal	\leftrightarrow +3 10 s 2x		 +3 8 s 1x 10 s \leftrightarrow +3		
Movimento pendular	\leftrightarrow +3 10 s 2x	 +3 8 s 1x 10 s \leftrightarrow +3			

Tabela 5.3: Resultados do segundo sujeito (cego, formação acadêmica - 6º ano, 49 anos de idade) do segundo protocolo experimental. Estrutura da tabela e os dados nela presente seguem as mesmas convenções que a tabela 5.1. [Ver texto para explicação detalhada.]











Tipo de movimento	(CM)				
Estático		 +2 10 s 2x	 ? 10 s 3x	 +3 10 s 3x	 ? 10 s 3x
Velocidade constante	$\rightarrow V_x = 1, V_y = 0$ +3 $\downarrow V_x = 0, V_y = 1$ +3 $\swarrow V_x = 1, V_y = 1$ +3	 2 : 2  +3 ?			
Acelerado					
Circular					
Oscilação vertical	\updownarrow +3 10 s 2x				
Oscilação horizontal	\leftrightarrow +3 10 s 2x				
Movimento pendular	\leftrightarrow +3 10 s 2x				

Tabela 5.4: Resultados do terceiro sujeito (amblíope, formação acadêmica - 9º ano, 50 anos de idade) do segundo protocolo experimental. Estrutura da tabela e os dados nela presente seguem as mesmas convenções que a tabela 5.1. [Ver texto para explicação detalhada.]

respostas, sendo uma possível explicação o facto desse sujeito ter um nível de formação académica mais avançada.

Capítulo 6

Conclusões

Nós desenvolvemos um modelo e uma aplicação para conversão automática de imagens e animações simples para som (sonificação), adequado para traduzir informação visual para cegos ou utilizadores com deficiências visuais. O som reproduzido dá informação acerca da forma e localização do objecto. Isto é realizado pela utilização de uma solução de duplo referencial e modulação do balanço e frequência do som para traduzir a informação acerca das posições de pontos geométricos. A informação da forma é dada pela movimentação do cursor à volta das arestas de um objecto e convertendo a localização do cursor (num referencial com origem no centro-de-massa do objecto) para som. A informação da posição/movimento do objecto é dada pela conversão da localização do centro-de-massa do objecto (em relação a um referencial fixo) para som. A informação sonora sobre a forma é produzida sequencialmente no tempo com a informação de posição/movimento.

Os resultados de testes de usabilidade com vários sujeitos cegos e amblíopes, sugerem que a solução pode fornecer informação facilmente reconhecida acerca da forma e movimento do objecto, pelo menos para formas geométricas e movimentos simples.

Em trabalho futuro pretende-se testar a solução com um maior número de sujeitos com deficiências visuais e experimentando com tipos adicionais de estímulos. Incluem-se aqui animações simples com conteúdo narrativo e imagens icónicas representando simbologia variada. Será pretendido que os sujeitos sejam capazes de apreender o conteúdo narrativo das sonificações. Duas mini-narrativas já preparadas inclui o cair de uma bola/projectil sobre a acção da gravidade com o amortecimento do

embate ao solo (figura 6.1) e a simulação de um foguete espacial numa trajectória parabólica ascendente (figura 6.2). Para as figuras icônicas pretende-se que os sujeitos sejam capazes de discriminar entre duas classes de estímulos (e.g. silhuetas de figuras masculinas e femininas ou silhuetas de formas vegetais e animais). Ver figura 6.3.

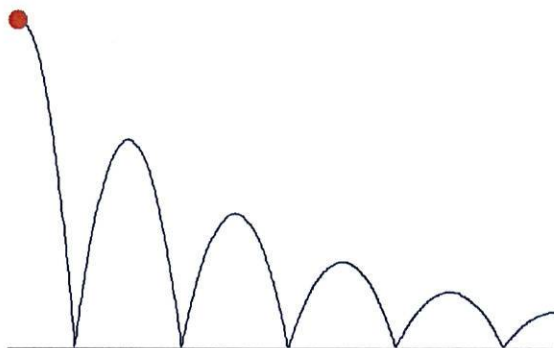


Figura 6.1: Mini-narrativa de uma bola sobre a acção da gravidade com o amortecimento do embate ao solo.

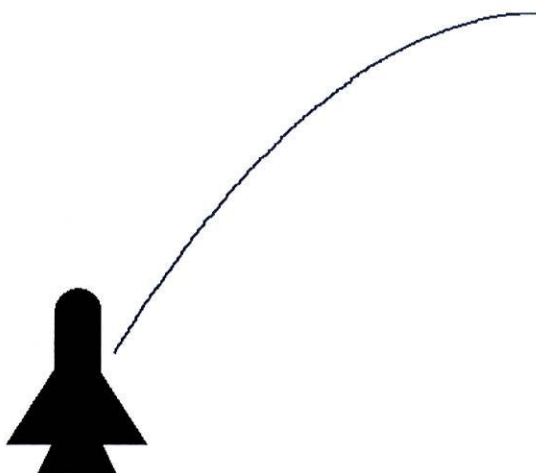


Figura 6.2: Mini-narrativa de uma simulação de um foguete espacial numa trajectória parabólica ascendente.

Pretende-se também vir a modificar/estender a solução de sonificação para transmitir informação adicional aos sujeitos. Por exemplo incorporar as sugestões feitas por um dos sujeitos entrevistados do segundo protocolo experimental, envolvendo a transmissão de informação sobre arestas de objectos e linhas curvilíneas. Este sujeito sugeriu também a criação de um versão interactiva da aplicação protótipo, para que os sujeitos pudessem experimentar e explorar o espaço de sonificações possíveis.

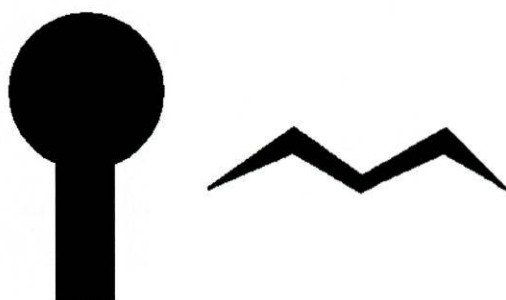


Figura 6.3: Silhuetas de formas vegetais e animais: **esquerda**) silhueta de uma árvore; **direita**) silhueta de uma gaivota.

Devera-se também fazer explorações adicionais nos valores mais apropriados para os parâmetros do algoritmo de sonificação (e.g. banda de frequências, timbre do som, etc.).

Uma vez construída uma solução robusta para o problema da sonificação, pretende-se também desenvolver algumas aplicações práticas identificando-se aqui alguns exemplos:

- Implementação de uma versão *web-ware* (usando as tecnologias/linguagens java ou javascript) da solução proposta para permitir a sonificação de imagens contidas em páginas de hipertexto.
- Estender um *browser* web (ou desenvolver um de raiz) que suporte a sonificação de imagens com ligação a páginas (X)HTML sem requerer que as páginas sejam pré-preparadas para utilizadores cegos.
- Aplicação da solução para sonificar dados estatísticos representados presentes em ficheiros de uma variedade de formatos.
- Criação de conteúdos e-learning de matemática e física para cegos recorrendo ao módulo de sonificação (entre outros de suporte para acessibilidade).
- Desenvolver uma ferramenta de desenho vectorial para cegos.
- Desenvolver uma ferramenta para criação de filmes de animação simples para utilizadores cegos, que poderá ser utilizado por utilizadores cegos e não cegos. Um nicho de aplicação será a criação de animações com histórias para crianças

cegas. Isto poderá envolver a inclusão de outras estruturas sonoras para aumentar a expressividade do realizador (e.g. produção sincronizada de *waves* de audio).

Comparando os resultados entre os diferentes sujeitos avaliados, foi possível identificar heterogeneidade no sucesso das respostas pelo que, se espera que diferentes utilizadores possam vir a ter mais ou menos dificuldade na utilização das aplicações a desenvolver. A aparente capacidade dos sujeitos/utilizadores ganharem familiaridade com os estímulos sonificados e a eventual disponibilização de uma ferramenta interactiva, deixa aberto a possibilidade das aplicações a desenvolver sejam de grande utilidade.

Referências

- [JSaS] Pedro Campos Jorge Simão and Pedro B. Silva. Sonification of form and movement for visual-impaired users. DSAI 2007, Software Development for Enhancing Accessibility and Fighting Info-exclusion.
- [KL] Hesham M. Kamel and James A. Landay. Constructing moving pictures eyes-free: An animation tool for the blind. CHI 2002, ACM Press.
- [KRS] Hesham M. Kamel, Patrick Roth, and Rashmi R. Sinha. Graphics and user's exploration via simple sonics (guess): Providing interrelational representation of objects in a non-visual environment. Proc. Int. Conference on Auditory Display, 2001, <http://www.acoustics.hut.fi/icad2001/>.
- [KSFL87] Rafael C. Gonzales K. S. Fu and C. S. G. Lee. *Robotics: Control, Sensing, Vision, and Intelligence*. McGraw-Hill Companies, 1987.
- [Mei] Peter Meijer. The vOICE - seeing with sound for the blind. <http://www.seeingwithsound.com>.
- [Sue] Jessie Suen. Conversion of images into sound. <http://www.its.caltech.edu/~chia/EE107/>.
- [vDD] Kees van Den Doel. Soundview: Sensing color images by kinesthetic audio. Proc. Int. Conference on Auditory Display, 2003.