



**João Gaspar Valente**

Licenciatura em Engenharia Informática

## **Sistema Multimodal para Captura e Anotação de Vídeo**

Dissertação para obtenção do Grau de Mestre em  
Engenharia Informática

Orientador: Prof. Doutor Nuno Manuel Robalo Correia,  
Professor Associado com Agregação, FCT/UNL

Júri:

Presidente: Prof. Doutor José Alberto Cardoso e Cunha  
Arguente: Prof. Doutora Maria Teresa Cacirola Chambel  
Vogal: Prof. Doutor Nuno Manuel Robalo Correia





**João Gaspar Valente**

Licenciatura em Engenharia Informática

## **Sistema Multimodal para Captura e Anotação de Vídeo**

Dissertação para obtenção do Grau de Mestre em  
Engenharia Informática

Orientador: Prof. Doutor Nuno Manuel Robalo Correia,  
Professor Associado com Agregação, FCT/UNL

Júri:

Presidente: Prof. Doutor José Alberto Cardoso e Cunha  
Arguente: Prof. Doutora Maria Teresa Caciro Chambel  
Vogal: Prof. Doutor Nuno Manuel Robalo Correia



## Sistema Multimodal para Captura e Anotação de Vídeo

### *Copyright*

*A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objectivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.*



## Agradecimentos

Agradecimentos são devidos a várias pessoas e entidades que possibilitaram a realização desta dissertação. Este trabalho, que, pessoalmente, marca o fim de uma etapa e dá início a uma nova fase da minha vida, não teria sido possível sem o apoio delas.

Em primeiro lugar, gostaria de agradecer ao Prof. Nuno Correia pela oportunidade que me ofereceu de explorar este tema interessante que é a Anotação de Vídeo e pela disponibilidade, apoio e orientação que me prestou durante o desenvolvimento desta dissertação. Gostaria de destacar o incentivo que me deu para chegar o mais longe possível, nomeadamente, para realizar tarefas que extrapolam a escrita da dissertação, como a candidatura a prémios ou a escrita de artigos científicos. Em segundo lugar, gostaria de agradecer o apoio mais técnico concedido incondicionalmente pelo doutorando Diogo Cabral, que sugeriu novas ideias para implementação e novas abordagens para os problemas encontrados. Com ele aprendi importantes noções de investigação científica e de investigação. Gostaria também de agradecer à *designer* Urândia Carvalho pela sua forte contribuição no desenho da interface para este projecto e ao mestrando João Silva pela ajuda na integração do seu trabalho neste projecto. Agradecimentos adicionais vão para a coordenadora do projecto Carla Fernandes e restantes responsáveis indirectamente relacionados e para os coreógrafos Stephan Jürgens e Rui Horta que, juntamente com os alunos de dança, se disponibilizaram a avaliar este projecto.

Especialmente, gostaria de agradecer à minha família e amigos, em particular aos meus pais e ao meu irmão, que ao longo dos anos têm sido um apoio essencial para que continue motivado e empenhado no caminho que tracei, e pela sua paciência para conversar sobre assuntos fora das suas áreas de estudo, como este trabalho.

Por fim, gostaria de agradecer ao Departamento de Informática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT/UNL) que, através do Centro de Informática e Tecnologias de Informação (CITI), me facultou as tecnologias necessárias para a implementação do projecto e me apoiou financeiramente o Mestrado em forma de Bolsas de Investigação.





## Resumo

---

Actualmente, existe um grande crescimento de popularidade de dispositivos tácteis e multi-toque e, por isso, é necessário cada vez mais o desenvolvimento de aplicações e interfaces que sejam elaboradas especificamente para esse tipo de interacção. Através da caneta digital e do toque humano, esta interacção torna-se natural para o utilizador, pois assemelha-se à interacção humano-papel-caneta. Este tipo de interacção pode ser aplicado a documentos multimédia, em particular ao controlo e à anotação de vídeo digital. Assim como as anotações em publicações impressas promovem a leitura activa, as anotações de vídeo promovem a visualização activa, facilitando a reflexão e a aprendizagem com o enriquecimento do conteúdo do vídeo. As formas de arte performativas, como a dança, são domínios que envolvem várias iterações de ensaio, onde a anotação de vídeo pode melhorar significativamente o processo criativo dos autores.

Assim, a solução proposta tem como objectivo o desenvolvimento de uma ferramenta que suporte a captura e anotação multimodal de vídeo em tempo real, a ser executada num Tablet PC, explorando a interacção bimanual de toque humano com a caneta digital. A ferramenta permite ao utilizador capturar um vídeo através de uma fonte e, simultaneamente, anotá-lo, de forma a documentar segmentos desse vídeo para futura recuperação e pesquisa e aumentar o seu conteúdo, associando-lhe um significado próprio. No âmbito da dança contemporânea, um coreógrafo pode utilizar o sistema a fim de registar e anotar um ensaio ou uma performance ao vivo, actuando como um bloco de notas digital que pode ser revisto e partilhado com os artistas.

Esta proposta está enquadrada no projecto TKB, desenvolvido em colaboração com a Faculdade de Ciências Sociais Humanas da Universidade Nova de Lisboa – FCSH/UNL, que inclui nos objectivos o desenvolvimento de um anotador de vídeo aplicado à dança contemporânea.

**Palavras-chave:** Anotação de Vídeo, Captura de Vídeo, Interface Baseada na Caneta Digital e Toque, Dispositivos Tácteis e Multi-toque, Dança Contemporânea.

---



## Abstract

---

Currently, there is a large growth in popularity of tactile and multi-touch devices and therefore it is increasingly necessary to develop applications and interfaces that explore this type of interaction. Through a stylus and touch, this interaction becomes natural to the user, because it resembles the human-paper-pen interaction. This type of interaction can be applied to multimedia documents, in particular to the control and annotation of digital video. As annotations in printed publications promote active reading, video annotations promote active viewing, favoring reflection and learning with enrichment of the video content. The performing art forms, such as dance, are all domains that involve several rehearsal iterations, where the video annotation can significantly improve the creative process of authors.

Therefore, the proposed solution aims at developing a tool that supports the capture and multimodal real-time annotation of video, running on a Tablet PC, which explores the bimanual interaction of human touch with a stylus. The tool allows the user to capture a video through a video source, and, simultaneously, annotate it, in order to document video segments for future research and retrieval and augment video content, associating it a personal meaning. In the context of contemporary dance, a choreographer can use the system to record and annotate a rehearsal or a live performance, becoming a digital notebook that can be reviewed and shared with the artists.

This proposal is done in the scope of the TKB project, developed in collaboration with Faculdade de Ciências Sociais Humanas da Universidade Nova de Lisboa- FCSH / UNL, that includes as one of the goals the development of a video annotator applied to contemporary dance.

**Keywords:** Video Annotation, Video Capture, Multi-touch Pen Based Interface, Tactile and Multi-touch Devices, Contemporary Dance.

---



# Índice

---

<b>1. Introdução</b>	<b>23</b>
1.1 Motivação .....	23
1.2 Descrição e Contexto.....	24
1.3 Solução Apresentada .....	25
1.4 Principais Contribuições.....	26
1.5 Estrutura do Relatório.....	27
<b>2. Trabalho Relacionado</b>	<b>29</b>
2.1 Anotação de Vídeo .....	29
2.1.1 Anotadores Clássicos .....	30
2.1.2 ELAN .....	31
2.2 Anotação de Vídeos de Dança.....	32
2.2.1 DanVideo .....	33
2.2.2 VideoTraces .....	33
2.3 Anotação de Vídeo Baseada na Caneta Digital .....	35
2.3.1 Durante a Captura do Vídeo.....	36
2.3.1.1 Marquee	36
2.3.1.2 M4Note	38
2.3.1.3 NoteLook	40
2.3.2 Após a Captura do Vídeo .....	41
2.3.2.1 LEAN	41
2.3.2.2 Ambulant Annotator	43
2.3.2.3 Videotater	44
2.4 Interação com Caneta Digital e Multi-Toque .....	46
2.4.1 Benefits of Bimanual Pen and Direct-Touch Interaction on Horizontal Interfaces .....	47
2.4.2 Manual Deskterity.....	48

<b>3.</b>	<b>Captura e Anotação de Vídeo</b>	<b>51</b>
3.1	Conceito.....	51
3.2	Requisitos .....	52
3.2.1	Requisitos de Desenho .....	52
3.2.2	Requisitos Funcionais .....	53
3.3	Solução .....	54
3.3.1	Tecnologias .....	55
3.3.2	Interface.....	56
3.3.2.1	Interface do Tablet PC	56
3.3.2.2	Interface do Dispositivo Móvel	58
3.3.3	Funcionalidades.....	59
3.3.3.1	Captura de Vídeo	59
3.3.3.2	Modos de Anotação de Vídeo	61
3.3.3.3	Tipos de Anotação de Vídeo	64
3.3.3.4	Navegação no Vídeo e Anotações	70
3.3.3.5	Edição e Visualização de Anotações	70
3.3.3.6	Gestão do Projecto	71
<b>4.</b>	<b>Realização</b>	<b>73</b>
4.1	Arquitectura.....	73
4.2	Captura de Vídeo.....	77
4.2.1	Abordagens Possíveis.....	78
4.2.2	Modos de Anotação de Vídeo .....	79
4.3	Captura de Áudio.....	82
4.4	Anotações de Vídeo.....	82
4.4.1	Tipos de Anotação de Vídeo .....	83
4.4.2	<i>Motion Tracking</i> .....	86
4.5	Armazenamento de Informação Gerada .....	86
4.6	Controlo Remoto .....	90
<b>5.</b>	<b>Avaliação</b>	<b>93</b>
5.1	Testes Preliminares.....	93
5.1.1	Descrição.....	94
5.1.2	Análise de Observações .....	94
5.2	Teste Final .....	96
5.2.1	Descrição.....	96

5.2.2	Inquérito aos Utilizadores .....	97
5.2.3	Análise de Resultados .....	97
5.2.3.1	Utilizador .....	98
5.2.3.2	Cenários de Utilização .....	98
5.2.3.3	Tipos de Anotação .....	99
5.2.3.4	Modos de Anotação .....	100
5.2.3.5	Experiência .....	100
<b>6.</b>	<b>Conclusões e Trabalho Futuro</b> .....	<b>103</b>
6.1	Conclusões.....	103
6.2	Trabalho Futuro .....	104
	<b>Bibliografia</b> .....	<b>105</b>
	<b>Anexo A – XML Schema</b> .....	<b>109</b>
	<b>Anexo B – Questionário</b> .....	<b>113</b>
	Enunciado .....	114
	Resultados.....	120

---





## Índice de Figuras

---

2.1 Interface do sistema ELAN [3].	32
2.2 Interface da adição/edição de micro anotações do sistema DanVideo [13].	33
2.3 Interface do sistema VideoTraces [12].	34
2.4 Partilha e comentários de <i>traces</i> [15].	35
2.5 Interface gráfica do Marquee [17].	37
2.6 M4Note Interface [1].	39
2.7 Anotação numa imagem capturada e transmitida do servidor para este Tablet no sistema NoteLook [19].	40
2.8 Controlo TLslider [20].	42
2.9 Controlo TLslider em forma de S [20].	42
2.10 Interface do sistema Ambulant Annotator [5].	44
2.11 Interface do sistema Videotater [22].	45
2.12 a) 5 frames anteriores à in frame; b) 5 frames posteriores à in frame; c) segmento seleccionado [22].	45
2.13 Desenho em formas livres (esquerda); Desenho com linhas rectas quando existe toque (direita) [23].	47
2.14 Escolha de cor bimanual [23].	47
2.15 Corte ao longo da régua (a); A imagem é cortada em dois pelo corte desenhado (b) [24].	48
2.16 Três sequências possíveis (sequencial, sobreposto e simultâneo) que causam comandos diferentes [23].	50
3.1 Diferentes tecnologias de input da ferramenta.	56
3.2 Interface versão <i>light</i> da ferramenta no Tablet PC.	58
3.3 Interface do dispositivo móvel executada num iPod Touch 3G.	59
3.4 Visualização simultânea das duas <i>streams</i> de vídeo.	60
3.5 Visualização das duas <i>streams</i> de vídeo sobrepostas.	61
3.6 Modo contínuo - anotações desvanecem-se com o tempo.	62
3.7 Modo suspenso - anotações permanecem visíveis.	63
3.8 Inserção de anotação de tinta digital.	65

3.9	Inserção de anotação de texto no modo Tablet PC.	66
3.10	Inserção de anotação na forma de hiperligação.	67
3.11	Inserção de anotação em forma de marcas com a barra de marcas aplicáveis em baixo.	68
3.12	Inserção de anotação em forma de áudio.	69
3.13	<i>Timeline</i> e visualização de todos os tipos de anotação gráfica criadas.	70
3.14	Menu <i>New Mark</i> que permite criar novas marcas personalizadas.	72
3.15	Menu <i>Navigator</i> que contém todas as anotações criadas com os seus instantes iniciais e finais.	72
4.1	Arquitectura geral do sistema.	74
4.2	Qt Designer (área à esquerda) e respectivo código gerado (área à direita).	75
4.3	Árvore de herança de algumas classes Qt.	75
4.4	Correspondência de <i>signals</i> de um <i>QToolButton</i> a <i>slots</i> de uma <i>QGroupBox</i> .	76
4.5	Estrutura de uma aplicação em openFrameworks.	77
4.6	Exibição e gravação de <i>frames</i> ao longo do tempo no modo contínuo.	80
4.7	Exibição e gravação de <i>frames</i> ao longo do tempo no modo atrasado.	81
4.8	Resumo do diagrama de classes dos tipos de anotação.	83
4.9	Resumo do diagrama de classes relativo ao tipo de anotação tinta digital.	84
4.10	Exemplo do comportamento do desvanecimento dos traços de tinta, ilustrando um tempo limite $t$ de 5 segundos.	85
4.11	<i>Motion tracking</i> com recurso ao Microsoft Kinect.	86
5.1	Primeiro teste preliminar com Stephan Jürgens realizado no dia 29 de Março.	94
5.2	Coreógrafo Rui Horta, em Montemor-o-Novo, interagindo com a ferramenta com a projecção da experiência atrás.	97
5.3	Resultado da classificação de frequência do eventual uso dos cenários de utilização.	99
5.4	Resultado da classificação de frequência do eventual uso dos tipos de anotação em cada cenário de utilização.	99
5.5	Resultado da classificação de frequência do eventual uso dos modos de anotação.	100
5.6	Resultado da classificação emocional da experiência.	101
A.1	XML Schema para o armazenamento das marcas de sistema e personalizadas existentes para aplicação.	110
A.2	XML Schema para armazenamento das anotações do tipo marcas.	110
A.3	XML Schema para armazenamento das anotações do tipo tinta digital.	111
A.4	XML Schema para armazenamento das anotações do tipo áudio.	111
A.5	XML Schema para armazenamento das anotações do tipo hiperligações.	112
A.6	XML Schema para armazenamento das anotações do tipo texto.	112

## Índice de Tabelas

---

2.1	Vantagens e desvantagens do sistema Marquee.	38
2.2	Vantagens e desvantagens do sistema M4Note.	40
2.3	Vantagens e desvantagens do sistema NoteLook.	41
2.4	Vantagens e desvantagens do sistema LEAN.	43
2.5	Vantagens e desvantagens do sistema Ambulant Annotator.	44
2.6	Vantagens e desvantagens do sistema Videotater.	46
2.7	Vantagens e desvantagens no uso da caneta ou de toque como input [23].	49
3.1	Comparação entre os diferentes modos de anotação.	64
3.2	Comparação entre os diferentes tipos de anotação.	69
4.1	Vantagens e desvantagens de cada tipo de captura quanto à carga de processamento exigida verificada.	79



## Índice de Listagens

---

4.1 Declaração do método <i>connect</i> .	76
4.2 Excerto do código correspondente à captura de vídeo em modo contínuo.	80
4.3 Excerto do código correspondente à captura de vídeo em modo atrasado.	81
4.4 Excerto do código correspondente à captura de áudio.	82
4.5 Excerto do ficheiro XML que representa o armazenamento de uma anotação de tinta digital.	87
4.6 Excerto do ficheiro XML que representa o armazenamento de uma anotação de texto.	88
4.7 Excerto do ficheiro XML que representa o armazenamento de uma hiperligação.	89
4.8 Excerto do ficheiro XML que representa o armazenamento de uma anotação de áudio.	89
4.9 Excerto do ficheiro XML que representa o armazenamento de uma marca.	90
4.10 Excerto do código correspondente à recepção e tratamento de mensagens OSC provenientes do controlo remoto.	91



# 1. Introdução

Actualmente, existe uma tendência para alterar a forma como se interage com um computador. Os tradicionais periféricos como o rato e o teclado não são os mais adequados quando se pretende uma interacção mais directa e natural com um computador. Uma alternativa a estes periféricos é o uso do toque humano combinado com o toque por caneta. Os dispositivos que dão uso a esta interacção podem ser vistos como uma evolução nas interfaces pessoa-máquina, no sentido em que há uma libertação do modelo tradicional do teclado e rato. O desenho de formas livres e posterior reconhecimento de caligrafia, a execução de gestos e controlo da pressão através do uso da caneta digital e a navegação natural e fluida através do uso do toque humano são características que podem ser vantajosas em aplicações que suportam a actividade criativa.

## 1.1 Motivação

Recentemente, a popularidade de dispositivos tácteis e multi-toque tornam essencial o desenvolvimento de aplicações e interfaces que sejam pensadas especificamente para este tipo de interacções. Estudos recentes concluem que a interacção disponibilizada por estes dispositivos, combinando toque humano com toque por caneta, tem várias vantagens, como o facto de se assemelhar a uma interacção familiar humano-papel-caneta. Assim, estes dispositivos, como o Tablet PC, proporcionam uma aproximação ao “papel electrónico”, não só pela analogia do tamanho do ecrã com o tamanho de uma folha de papel, como também pelo seu poder de processamento. Por estas razões, este tipo de interacção pode ser aplicado à anotação de documentos multimédia, mais particularmente àqueles que envolvem vídeo digital.

A caneta, definida como “*this rapidly moving little servant of human intellect*” (S.R. Hirsch) ou “*the tongue of the mind*” (M. Cervantes Saavedra) é uma ferramenta antiga que resultou de uma necessidade de comunicar e memorizar informação. Esta informação pode integrar anotações escritas ou desenhadas em documentos. Uma anotação é uma forma de marcar um documento, com o objectivo de facilitar a interpretação e compreensão do seu conteúdo [1]. Uma anotação de vídeo,

além de ter esse objectivo, visa também a preparação da recuperação posterior do vídeo em causa. Com isto, as anotações, quer sejam multimédia (por voz ou vídeo), textuais ou em tinta, fornecem notas pessoais que definem contextos e significados em segmentos do documento, tornando o conteúdo mais interactivo e activo. Uma das áreas interessantes de anotação que tem sido alvo de vários estudos é a anotação de vídeos de dança.

Embora a dança tenha sido beneficiada pelos avanços de tecnologia na captura e armazenamento de informação digital, que permitiram a produção rápida de vídeos de dança, a anotação e consulta de vídeos de dança são ainda grandes desafios. Segundo a *Encyclopedia Britannica*, uma anotação de dança é a captura de movimentos de dança através do uso de símbolos escritos. É uma forma simbólica de representar os movimentos dos bailarinos. Um movimento da dança representa uma pose básica, gesto ou acção executada pelo bailarino. Uma coreografia de dança é um conjunto de movimentos de dança. Uma colecção de coreografias denota uma dança realizada pelo(s) bailarino(s). Os vídeos de dança contêm uma semântica rica, como as emoções dos bailarinos ou os passos de dança [2]. Estes passos são movimentos executados pelos bailarinos, os quais podem ser acompanhados automaticamente através de *motion tracking*. Estas anotações podem ser usadas para arquivar coreografias e para descrever o seu conteúdo, assistindo alunos de dança nos seus ensaios.

Num ensaio típico de coreografia, o coreógrafo dá *feedback* verbal aos bailarinos, mas esse *feedback* está dependente da memória que ele tem dos movimentos correctos e errados executados pelos bailarinos. Ao ser capturada e anotada, a coreografia pode ser visualizada múltiplas vezes, dando oportunidade ao coreógrafo para rever comentários que fez anteriormente, observar aspectos da coreografia que não tinha observado e focar comentários em elementos mais específicos da dança. Além disso, dá a possibilidade aos bailarinos de visualizarem a sua performance de um ponto de vista exterior. Com a anotação de vídeo de dança, há um meio de comunicação entre coreógrafos e bailarinos mais preciso, que gera uma discussão e reflexão sobre os eventos do ensaio. Por isso, os sistemas de anotação de vídeo de dança podem ser entendidos como sistemas para cooperação/colaboração, em que o objectivo é usualmente indicarem a alguém partes interessantes de um documento como um método de coordenação da actividade ou como um método de partilha de notas em tempo real numa situação de trabalho.

## 1.2 Descrição e Contexto

Esta dissertação está enquadrada no projecto TKB (Transmedia Knowledge Base) com a referência PTDC/EAT-AVP/098220/2008, que se encontra em desenvolvimento pelo Centro de Informática e Tecnologias de Informação (CITI) da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa em colaboração com a Faculdade de Ciências Sociais Humanas – FCSH/UNL e financiado por



fundos nacionais através da FCT/MCTES (PIDDAC). TKB é um projecto transdisciplinar, que visa o desenho e construção de uma base de conhecimento multimodal para documentar, anotar e suportar a criação de coreografias de dança contemporânea. O projecto procura proporcionar um espaço de pesquisa para a exploração rigorosa da relação entre a linguística, a dança, novos *media* digitais e pensamento/consciência. O projecto TKB implica a articulação de três componentes complementares, cada um com tarefas e objectivos específicos:

- 1) Anotação linguística, mais propriamente a anotação de vídeos de ensaios de dança com a criação de um glossário especializado de termos para a construção de um arquivo de dança contemporânea portuguesa.
- 2) Software para anotação e análise de movimento do corpo humano em sequências de vídeo de dança.
- 3) Ferramenta orientada ao processo de criação e arquivos individuais dos coreógrafos.

O componente 1) irá ser realizado na Faculdade de Ciências Sociais Humanas da Universidade Nova de Lisboa – FCSH/UNL. Os componentes 2) e 3) irão ser realizados na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa –FCT/UNL, sendo que o componente 2) irá ser realizado pelo mestrando João Silva e mais tarde integrado na solução proposta. O componente 3) é o principal objectivo desta dissertação.

Assim, esta dissertação visa a concepção de uma interface para suportar uma ferramenta orientada à criação, através da anotação de vídeo, com o propósito de apoiar o processo criativo coreográfico. Deverá ser uma ferramenta auxiliar para a tarefa do criador em coreografias em tempo real e para as suas intenções de arquivo.

Espera-se que o projecto TKB seja relevante para companhias de dança, escolas superiores de dança, museus de arte contemporâneos e em comunidades científicas de linguística cognitiva, *corpora* multimodal e novas tecnologias de *media*.

### **1.3 Solução Apresentada**

Com o intuito de apoiar o processo criativo de criação de composições de dança, a solução proposta tem assim por objectivo desenvolver um sistema de captura e anotação multimodal de vídeo em tempo real, com suporte para uma interface que combina a interacção de toque humano com toque por caneta, usando um Tablet PC. O sistema dá a possibilidade ao utilizador, neste caso o coreógrafo, de capturar um vídeo de um ensaio ou de uma performance ao vivo, através de uma câmara web ligada ou incorporada no Tablet PC, e anotá-lo em tempo real, de forma a documentar segmentos desse vídeo para futura recuperação e pesquisa e, com isso, treinar e preparar os participantes do

espectáculo de forma mais eficiente. Após o processo de captura, também é possível criar anotações ou editar e visualizar sincronizadamente as anotações anteriormente criadas durante a captura.

Para corresponder às necessidades dos coreógrafos, é permitido ao utilizador a criação de dois tipos de anotação: anotações regulares e marcas (*bookmarks*). As marcas consistem em conceitos, definidos pelo coreógrafo, que representam uma *keyword*, como “luz” ou “palco”. Em contraste, as anotações regulares não têm uma estrutura pré-definida. Estas podem ser do tipo texto, áudio ou baseadas na caneta digital e são introduzidas por um teclado virtual, microfone *Bluetooth* ou caneta digital respectivamente. As anotações baseadas na caneta digital e anotações por voz permitem tipos distintos de interações que não são suportadas pelo teclado. As anotações por voz fornecem um método instantâneo e desprovido de localização para criação de anotações. As anotações baseadas na caneta digital proporcionam a possibilidade de desenhar no conteúdo do vídeo, permitindo mais liberdade no processo criativo.

Como referido anteriormente, a tecnologia usada para desenvolver esta solução é o Tablet PC, um computador portátil que combina a interação entre caneta digital e multi-toque. Pelo facto de reproduzir uma interação humano-papel-caneta, que constitui uma interação natural e familiar para o utilizador no processo de anotação, pelo seu poder de processamento, tamanho de ecrã, peso reduzido e elevada portabilidade, esta é a tecnologia adequada para este sistema.

A solução proposta encontra-se desenvolvida na linguagem C++, recorrendo à plataforma *openFrameworks* 0.062<sup>1</sup>, uma biblioteca *open-source* desenhada para programação criativa e ao *toolkit Qt GUI* 4.7.3<sup>2</sup>, uma biblioteca multi-plataforma para desenvolver interfaces gráficas de utilizador. Adicionalmente, há a possibilidade de controlar remotamente o sistema a partir de um dispositivo móvel, através do protocolo de comunicação *Open Sound Control* (OSC)<sup>3</sup>. Para efeitos de arquivo de vídeos de dança, irá ser utilizado o sistema ELAN [3].

Apesar da solução proposta ser desenvolvida especificamente para a área de dança contemporânea, também pode ser aplicada noutras áreas que envolvam igualmente processos de criação, treino e preparação, como por exemplo os desportos de alta competição ou a educação.

## 1.4 Principais Contribuições

As principais contribuições atingidas com a implementação da solução apresentada são as seguintes:

---

<sup>1</sup> <http://www.openframeworks.cc/>

<sup>2</sup> <http://qt.nokia.com/>

<sup>3</sup> <http://opensoundcontrol.org/introduction-osc>

- **Interface Multi-Toque Baseada na Caneta Digital:** Desenvolvimento de uma interface que combina a interacção de toque por caneta com o multi-toque humano, que permitirá assistir o processo criativo da dança dos coreógrafos.
- **Anotação Multimodal de Vídeo:** Desenvolvimento dos componentes necessários para suportar a funcionalidade de anotação multimodal de vídeo. Inclui a criação, edição, armazenamento e visualização de anotações.
- **Captura de Duas Fontes de Vídeo Simultaneamente:** Suporte para captura até duas fontes de vídeo simultaneamente (multi-câmara). Gravação em disco dos vídeos resultantes.
- **Controlo Remoto de Funcionalidades Via Dispositivo Móvel:** Desenvolvimento de uma interface para um dispositivo móvel, com o intuito de aceder e controlar remotamente algumas funcionalidades do sistema.
- **Integração entre Qt GUI e openFrameworks:** Integração das duas diferentes plataformas, de forma a constituir um *framework* único para o desenvolvimento do sistema. A plataforma Qt GUI servirá como suporte para o desenho e implementação da interface e a plataforma openFrameworks servirá como suporte ao desenvolvimento das restantes funcionalidades do sistema.
- **Avaliação de Utilizadores:** Realização de testes de usabilidade do sistema, tanto da interface como da aplicação em si. Esta avaliação irá ser elaborada por peritos de dança contemporânea, incluindo os coreógrafos Rui Horta e Stephan Jürgens e alunos de dança. Seguir-se-á uma análise dos resultados.
- **Publicação de Artigos Científicos:** Publicação de dois artigos científicos sobre o propósito desta dissertação para duas conferências internacionais da área: ACM CHI 2011 WIP<sup>4</sup> e ACM Multimedia 2011<sup>5</sup>.

## 1.5 Estrutura do Relatório

O presente documento encontra-se estruturado em seis capítulos, descritos de seguida:

**Capítulo 1 - Introdução:** Apresenta uma visão geral da dissertação, no que diz respeito à motivação, ao contexto, à descrição do problema, à solução apresentada e às principais contribuições previstas.

**Capítulo 2 - Trabalho Relacionado:** Estabelece uma ligação entre os objectivos desta dissertação e o trabalho relacionado. São apresentados sistemas de anotação de vídeo geral, com ênfase no sistema ELAN, sistemas de anotação de vídeos de dança e sistemas de anotação baseados na caneta digital.

---

<sup>4</sup> <http://www.chi2011.org/>

<sup>5</sup> <http://www.acmmm11.org/>

Dentro dos sistemas de anotação baseados na caneta digital, são descritos sistemas que suportam a anotação durante e após a captura do vídeo, juntamente com as vantagens e desvantagens de cada sistema relativamente à solução proposta e com uma discussão geral de cada secção. Por fim, são apresentados sistemas que exploram a combinação da interacção com caneta digital e multi-toque e uma análise desta interacção.

**Capítulo 3 – Captura e Anotação de Vídeo:** Descreve a solução implementada, indicando inicialmente o conceito e requisitos a cumprir, seguido da apresentação da interface e funcionalidades que satisfazem os requisitos, bem como das tecnologias utilizadas para a sua realização.

**Capítulo 4 – Realização:** Explica como a solução foi realizada, apresentando a arquitectura do sistema e os detalhes mais relevantes na implementação das principais funcionalidades descritas anteriormente.

**Capítulo 5 – Avaliação:** Apresenta uma análise dos resultados obtidos da avaliação da solução realizada por alunos e peritos da dança contemporânea, através do preenchimento de um inquérito e observações registadas durante a interacção.

**Capítulo 6 – Conclusões e Trabalho Futuro:** Inclui uma apreciação do trabalho realizado, assim como sugestões para possíveis caminhos a tomar para a evolução do mesmo.

## 2. Trabalho Relacionado

Neste capítulo apresentam-se os conceitos e as técnicas necessárias para o desenvolvimento da solução proposta. Para tal, tornou-se fundamental rever um conjunto de trabalhos de investigação relacionados com o tema desta dissertação. Os tópicos serão abordados em quatro secções principais. Na primeira secção, são apresentados sistemas de anotação de vídeo em geral. Na segunda secção, são apresentados sistemas de anotação de vídeo elaborados especificamente para anotação de vídeos de dança. Na terceira secção, são apresentados sistemas de anotação de vídeo baseados na caneta digital, isto é, usando a caneta como dispositivo de entrada para escrever/desenhar anotações, que suportam anotação de vídeo durante a captura (em tempo real), tal como a solução proposta, e após a captura (*a posteriori*) do vídeo. Por último, são analisados trabalhos que exploram o conceito e técnicas de interacção com caneta digital e multi-toque, que irá ser a interacção usada na solução proposta.

### 2.1 Anotação de Vídeo

Do ponto de vista da interacção entre utilizador e documento, uma anotação é uma forma de marcar um documento, com o objectivo de facilitar a interpretação e compreensão do seu conteúdo [1]. De acordo com Bottoni et al. [4], a anotação suporta várias funções cognitivas: (1) Lembrança, através do destaque das partes mais importantes do documento anotado, (2) Raciocínio, através da adição de ideias, observações e questões próprias e a (3) Compreensão, através da transformação da informação contida no documento em representações verbais pessoais. Este autor refere também que uma nota digital é composta por dois componentes principais: meta dados e conteúdo. O primeiro componente refere-se a um conjunto de atributos como o autor, título, data de criação e modificação da nota e o segundo componente ao conteúdo multimédia da nota, que pode ser constituído por texto, imagem, vídeo ou áudio. Bulterman [5] sugeriu que existissem dois tipos de anotação para diferenciar os dois grandes usos na anotação de documentos. Esses tipos são as anotações hierárquicas e as anotações ao nível de pares (*peer level*). As anotações hierárquicas são anotações (incluindo os meta dados) que fornecem uma classificação abstracta do conteúdo para um dado uso ou ontologia. São representadas

como uma colecção de *strings*. As anotações *peer-level* providenciam informação adicional que resultam num conteúdo aumentado. Estas anotações podem ser tanto discretas (texto e imagens) como contínuas (áudio).

As anotações digitais têm claras vantagens em relação a anotações no papel, pois podem ser procuradas, partilhadas, arquivadas e facilmente manipuláveis [6]. Além disso, estas anotações constituem um novo documento, em vez de serem modificações a um documento existente, como acontece com as anotações em papel. Em relação à anotação de vídeo, além de ter uma dimensão espacial, como nas anotações em papel, tem também uma dimensão temporal que tem de ser considerada [7], ou seja, qual o intervalo de tempo no vídeo em que a anotação fica associada. Para o vídeo, a anotação é habitualmente uma forma de descrever o conteúdo para posterior recuperação [7].

Para a anotação de vídeo, existe uma norma chamada MPEG-7 [8], que facilita a gestão de dados multimédia, estabelecendo uma padronização na forma de descrição de diferentes aspectos de multimédia em diferentes níveis de abstracção, contribuindo fortemente para a utilização de meta dados, já que grande parte dos descritores MPEG-7 associados às informações semânticas são textuais.

Existem vários sistemas que exploram o conceito de anotação de vídeo, quer seja para enriquecer o seu conteúdo ou para facilitar o arquivo, pesquisa e reutilização de vídeos. Pode ser efectuada de forma manual ou recorrendo a algoritmos de processamento, mas actualmente as aproximações mais promissoras utilizam uma combinação destas duas técnicas. Segue-se um levantamento do estado da arte no que diz respeito à anotação de vídeo em geral, com destaque para o sistema ELAN, que irá ser usado para fins de arquivo dos vídeos.

### 2.1.1 Anotadores Clássicos

GALATEA [9] data de 1976 e é um dos sistemas de anotação de vídeo mais antigos. Neste sistema, imagens animadas ou gráficos podem ser sobrepostos directamente no vídeo. Os utilizadores têm a possibilidade de escrever directamente no vídeo usando uma caneta digital, dando ao sistema uma única e verdadeira interface de manipulação no tempo em que foi desenvolvido. O sistema EVA [10] consiste numa ferramenta de anotação com a capacidade de anotar segmentos de vídeo, com *keywords* em tempo real ou depois de um evento. As transcrições de texto neste sistema podiam aparecer como legendas sincronizadas com o vídeo. O sistema VANNA [11] suporta uma variedade de dispositivos de entrada para manipular anotação em tempo real e uma análise detalhada dos dados do vídeo. AntV (*Annotations in Video*) [7] é um anotador de vídeo que pode ser usado para adicionar ou editar anotações de texto, de imagem ou até de vídeo. Estas podem ser ligadas ao vídeo usando intervalos de tempo ou posições no espaço. VideoAnnEx [12] foi desenvolvido pela IBM e permite ao utilizador anotar sequências de vídeo com os meta-dados MPEG-7. As anotações são guardadas usando descritores MPEG-7 em ficheiros XML. A ferramenta também pode abrir ficheiros MPEG-7 para visualizar a anotação para o ficheiro de vídeo correspondente. MRAS (*Microsoft Research*

*Annotation System*) [13] é um protótipo de anotação de vídeo colaborativo, que apresenta uma interface baseada na Web para adicionar e visualizar anotações de texto ou áudio. Permite aos utilizadores a partilha de anotações através de um servidor.

### 2.1.2 ELAN

ELAN (EUDICO Annotator Linguistic) [3] é uma ferramenta de anotação que permite criar, editar, visualizar e pesquisar manualmente anotações textuais de vídeo e áudio. É destinado especificamente para a análise da linguagem, mas pode ser usado por todos que trabalham com vídeo e/ou áudio, para fins de análise, anotação e documentação. Suporta a associação de, no máximo, seis arquivos de vídeo com um documento de anotação único, o que pode ser útil para gravações multi-câmara. Quatro deles podem ser exibidos simultaneamente na janela principal. Através de um menu que lista todos os vídeos associados, é possível alterar a combinação de vídeos exibidos.

Neste sistema, as anotações são agrupadas em *tiers*. Um *tier* é uma camada definida pelo utilizador que contém anotações que partilham as mesmas restrições na sua estrutura, conteúdo e/ou características de alinhamento. Podem fazer parte de uma hierarquia de *tiers*, ou seja, cada *tier* pode ter um *tier* pai e/ou *tiers* dependentes.

As anotações são escritas manualmente pelo utilizador, através da definição prévia de um segmento na *timeline* onde irá ser associada, como representa a figura 2.1. Esta identificação de segmentos é realizada com o arrastamento do rato no intervalo de tempo pretendido. Ainda nessa figura, é possível visualizar a interface global do sistema, com o vídeo que está a ser anotado localizado no canto superior esquerdo e as notas correspondentes localizadas no lado direito em forma de grelha. Existem diversos visualizadores de anotações como este para as anotações que foram criadas. Por baixo do vídeo, encontram-se os controlos para a reprodução do vídeo, que além de terem as funções comuns de começar, parar, ir para o início ou fim da reprodução, também dão a possibilidade de navegar no vídeo com diferentes factores de escala, como por *frame*, por segundo ou por página da *timeline*. Existem também controlos para navegar para a próxima ou anterior anotação criada. Os *tiers* estão presentes no lado esquerdo da *timeline*.

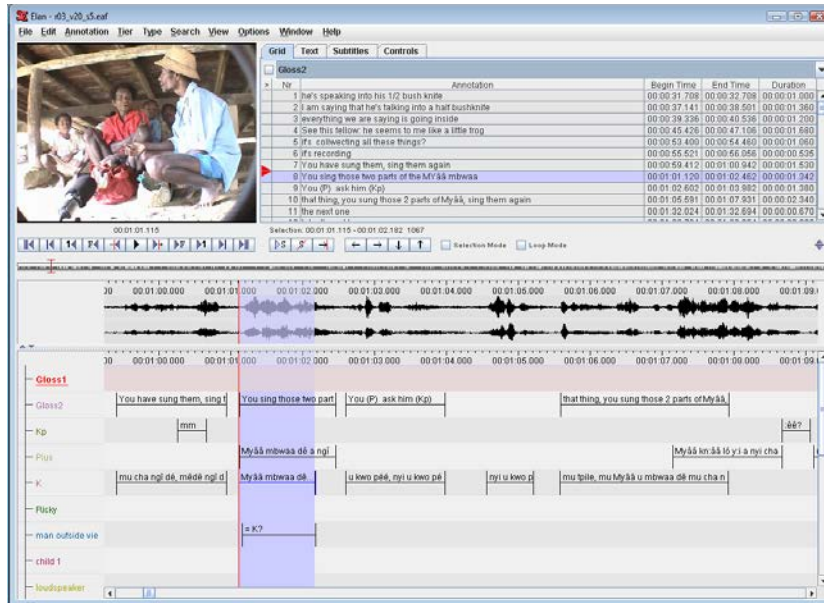


Figura 2.1: Interface do sistema ELAN [3].

O utilizador também tem a possibilidade de construir um *Controlled Vocabulary*, que é a definição de uma lista de anotações, que podem ser aplicadas posteriormente no momento da criação ou edição de anotações.

Também suporta a pesquisa por anotações, recorrendo a consultas complexas que combinam restrições temporais com restrições estruturais. Cada consulta pode ser baseada na estrutura de anotações, isto é, na relação entre *tiers* e entre anotações do mesmo *tier* ou de *tiers* relacionados, ou em relações temporais entre anotações, mais especificamente, na relação da posição da anotação numa sequência de anotações. As consultas efectuadas podem ser gravadas ou exportadas.

Existe ainda a opção de obter estatísticas sobre anotações do *tier* seleccionado ou de todos os *tiers*, como a ocorrência, frequência e duração média de cada anotação.

## 2.2 Anotação de Vídeos de Dança

Antigamente, os bailarinos passavam o conhecimento da dança verbalmente para as gerações seguintes [14]. No entanto, esse conhecimento estava limitado à memória dos bailarinos e muitas produções de dança foram perdidas devido a esse facto. Tradicionalmente, as anotações de dança são usadas para arquivar coreografias e ressuscitar produções de dança pelos bailarinos e coreógrafos [14]. Além desse uso, também podem servir para comunicação entre bailarinos e coreógrafos num ambiente de aprendizagem e reflexão [15].

Existem sistemas de anotação de vídeo elaborados especificamente para vídeos de dança, já que um vídeo de dança contém uma semântica rica, como os passos de dança ou as emoções dos bailarinos [2]. A dança contém uma componente visual (passos de dança), de áudio (música) e de texto (letras da música), o que constitui um domínio interessante da área de multimédia [2].



### 2.2.1 DanVideo

DanVideo [2] é um sistema para anotar manualmente as macro e micro características dos vídeos de dança por profissionais da dança. Gera uma instância MPEG-7 semi-automaticamente em formato XML, que está em conformidade com o esquema MPEG-7, produzindo metadados das anotações de dança criadas. Tem também um motor de busca que suporta consultas dos utilizadores (coreógrafos e alunos de dança) e devolve semântica da dança a partir dos metadados. Utiliza uma ontologia de domínio específico para processar consultas baseada no conhecimento. Este sistema pode ser utilizado para apoio ao treino e aprendizagem da dança.

O utilizador anota as macro e micro características manualmente, visualizando um vídeo. As macro anotações representam as características abstractas do vídeo como os detalhes dos bailarinos (ID, nome, telefone, email), música, origem da dança, tipo de dança, contexto (ensaio, competição, etc.), data e tempo de gravação, tipo do local da performance (teatro, praia, etc.), entre outros. As micro anotações dizem respeito a características específicas dos passos de dança como os actores (papéis) do evento, agentes dos actores e conceitos (emoções) reveladas pelos actores. Também descreve as várias relações entre actores, eventos, agentes e conceitos como as relações espaciais, temporais, semânticas e de movimento. A interface para inserir ou editar micro anotações está presente na figura 2.2.

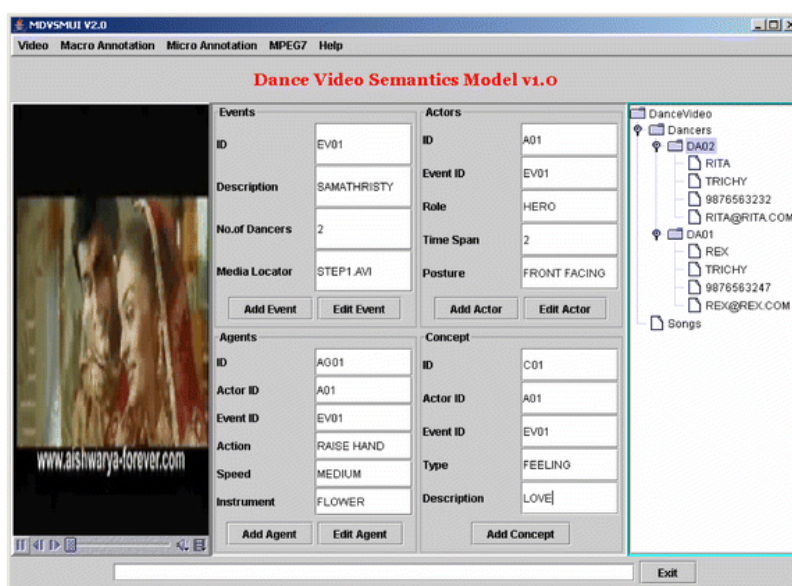


Figura 2.2: Interface da adição/edição de micro anotações do sistema DanVideo [13].

### 2.2.2 VideoTraces

VideoTraces [15] é uma ferramenta de anotação de vídeo que pode ser usada para facilitar a reflexão e avaliação crítica na dança. Permite aos utilizadores capturarem o vídeo e anotá-lo por gestos ou por voz. As *video traces* (anotações) resultantes são gravadas e podem ser visualizadas, partilhadas e

comentadas pelo criador e outros, fornecendo aos instrutores e alunos de dança uma forma de ver e reflectir sobre eventos que normalmente são efémeros.

Este sistema proporciona uma interface que permite ao utilizador capturar, gravar e organizar vídeos. Os utilizadores podem capturar vídeos directamente de uma câmara ligada ao computador ou podem usar um vídeo já capturado. Para anotar um vídeo, os utilizadores visualizam-no, falando sobre o que vêem. Movendo o cursor do rato para a *frame* de vídeo, é disponibilizada uma *finger tool* (ferramenta de dedo), que permite apontar para diferentes partes do vídeo enquanto é reproduzido. Também existem opções de reprodução do vídeo, como o controlo da velocidade, a suspensão e retrocesso da reprodução para anotar um ponto com maior detalhe ou para elaborar um segundo comentário numa secção particular respectivamente. O áudio, gestos (movimento com o cursor) e mudança na reprodução do vídeo são gravados e preservados na *video trace* resultante.



**Figura 2.3:** Interface do sistema VideoTraces [12]. Em baixo na secção “Audio”, existe um botão para inserir uma anotação de áudio (“Record Audio”). Por baixo, existem as opções de controlo do vídeo (reprodução regular, retroceder, pausar, reprodução lenta e reprodução rápida).

Para a discussão de *traces*, estas podem ser ligadas a outras de outros utilizadores, isto é, os utilizadores podem elaborar uma *trace* em resposta a outra *trace* de outro utilizador. Estas respostas aparecem visualmente ligadas e associadas com comentários, como ilustra a figura 2.4:



**Figura 2.4:** Partilha e comentários de *traces* [15]. Escolhendo a opção *New Topic*, é criada uma *trace*; Escolhendo a opção *Respond to Trace*, é criada uma discussão das *traces* no respectivo tópico; No canto esquerdo, está representado a primeira *frame* do vídeo associado à *trace* escolhida.

Para testar este sistema, os autores pediram ao instrutor de dança que o usasse, anotando uma coreografia (*trace*) num ensaio dos seus alunos. Os alunos depois foram convidados a visualizarem essas anotações e responderem aos comentários do instrutor. No final, os alunos relataram que visualizar as anotações do instrutor sobre o seu trabalho foi o componente do curso que mais contribuiu para a sua aprendizagem. Indicaram que consideraram o sistema útil para poderem ver o seu ensaio e ouvir comentários múltiplas vezes. Além disso, tiveram a oportunidade de ver o seu trabalho através de uma vista diferente. O instrutor referiu que este sistema encorajou os alunos a reflectir de forma mais crítica sobre os seus trabalhos e os trabalhos dos outros.

## 2.3 Anotação de Vídeo Baseada na Caneta Digital

O modelo de tinta e caneta digital dos Tablet PC é umas das várias aproximações para anotar objectos multimédia [5]. A combinação entre vídeo digital e a tecnologia baseada na caneta digital pode ser uma aproximação do “papel electrónico”, não só pelo facto de um Tablet PC ter um tamanho de ecrã semelhante ao tamanho de um papel A4, mas também pela possibilidade de se associar notas escritas ou desenhadas a segmentos de vídeo [16].

As anotações de vídeo baseadas na caneta digital, além de terem as mesmas restrições de espaço e tempo que as anotações de vídeo, têm propriedades diferentes e únicas. Por exemplo, cada anotação deste tipo pode ser considerada como sendo um conjunto de traços de tinta, que tem vários atributos como a espessura, cor ou tipo de tracejado que as diferenciam umas das outras. Assim, existe a possibilidade de desenhar traços com a caneta consideradas anotações, que oferecem ao utilizador total liberdade na sua elaboração, de forma rápida e prática. Estas anotações tanto podem ser texto escrito à mão como desenhos livres. Existem vários sistemas que exploram este conceito, para funcionarem especificamente com este tipo de interação e tecnologia associada. Existem sistemas

que suportam a anotação durante a captura do vídeo e outros que somente suportam após a captura do vídeo.

### 2.3.1 Durante a Captura do Vídeo

Os sistemas apresentados nesta secção são sistemas de anotação de vídeo que suportam a anotação durante a captura do vídeo, isto é, em tempo real. Assim, existe a possibilidade de capturar um vídeo e anotá-lo ao mesmo tempo. Contrariamente aos sistemas que somente suportam a anotação após a captura do vídeo, estes sistemas têm a restrição do tempo associado, já que é necessário efectuar anotações num determinado intervalo de tempo que corresponde ao tempo de captura do evento. Existem outros problemas a resolver nestes sistemas, como por exemplo qual o instante inicial e final de uma anotação feita em tempo real ou qual o comportamento da visualização da captura do vídeo no momento em que se insere a anotação. Estes problemas tornam-se mais facilitados em sistemas após a captura, onde existe um controlo total da reprodução do vídeo.

A anotação em tempo real fornece uma estrutura inicial para recuperação [17]. Isto é especialmente necessário para utilizadores que desejam uma descrição mais completa do material capturado, já que a primeira passagem de anotações lhes proporciona uma visão geral do mesmo [17]. Estes sistemas permitem sempre estender ou editar as anotações após a captura do vídeo, mas a característica que os define é a anotação em tempo real.

#### 2.3.1.1 Marquee

Marquee [17] é uma ferramenta de vídeo para *logging*/anotação, baseada na interacção com caneta digital, que permite aos utilizadores correlacionar notas e *keywords* com um vídeo durante a sua captura, isto é, em tempo real. O método de design desta ferramenta foi composto por três fases: (1) a criação de protótipos *Hypercard* para explorar diferentes desenhos de interface, (2) a utilização de protótipos em papel para testar as assumpções anteriores e finalmente (3) a criação de um protótipo funcional final para avaliar informalmente o sistema.

Um dos problemas a resolver neste sistema era a coordenação em tempo real da anotação livre com a aplicação de *keywords*. Inicialmente, na interface existia uma área para anotação e outra área com *keywords* para serem aplicadas. As *keywords* eram botões que se podiam ligar ou desligar em tempo real associando as mesmas ao início e fim de eventos particulares no vídeo respectivamente. Depois de vários testes de usabilidade efectuados a esta interface, os autores chegaram à conclusão que os utilizadores não conseguiram coordenar estas duas operações ao mesmo tempo, ou seja, ou anotavam livremente com a caneta ou aplicavam *keywords*.

Estes protótipos e respectivos testes iniciais deram origem a um protótipo final que tentava resolver os problemas iniciais. Marquee executa-se num tablet *Wacom* que está ligado a um computador *Macintosh*. O computador, por sua vez, está ligado a um dispositivo de captura de vídeo. Para anotar, aplicar *keywords* e controlar o vídeo, usa-se a caneta digital como dispositivo de entrada. A interface é composta por três áreas: *keyword palette*, que contém as *keywords* a serem aplicadas, *keyword stripping area*, que associa as *keywords* a *timezones* e *note taking area*, que representa a área onde o utilizador cria *timezones* e as anotações respectivas.

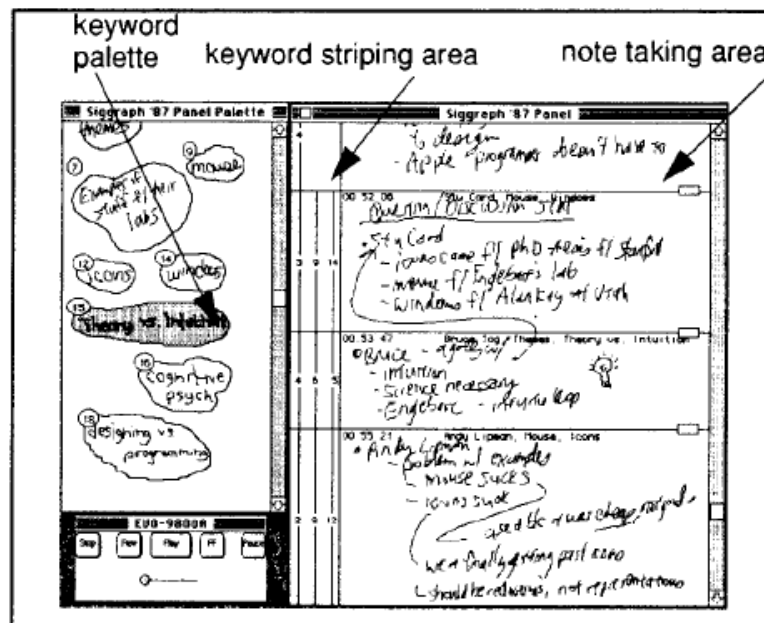


Figura 2.5: Interface gráfica do Marquee [17].

Para escrever notas em tempo real, o utilizador usa a *note taking area*. Para isso, cria *timezones*, que definem segmentos de vídeo, desenhando uma linha horizontal nesta área. A partir desse momento, o segmento está definido no intervalo de tempo entre a linha anterior e a linha desenhada e as notas que forem escritas/desenhadas neste segmento são associadas automaticamente. Para associar *keywords*, o utilizador selecciona uma *keyword* na *keyword palette* e desenha uma linha vertical na *keyword stripping area*. Os segmentos correspondentes onde a linha for desenhada ficam associados com a *keyword*, através da representação do seu identificador. Com este método de associar *keywords*, ficou resolvido o problema inicial, já que não é preciso ligar e desligar botões quando os eventos começam ou terminam, mas sim desenhando uma linha vertical de uma só vez que define o início e o fim de um evento.

Para reutilização e procura, usam-se as *keywords* traduzidas para texto legível para a máquina. Esta tradução é feita pelo utilizador manualmente. Também é possível criar novas *keywords*, desenhando um círculo à volta da nota desejada.

Elaboraram-se testes para este protótipo e os utilizadores na sua maioria concordaram que, não só Marquee é uma ferramenta poderosa de *logging* em tempo real, como também de recuperação desses *logs*.

**Tabela 2.1:** Vantagens e desvantagens do sistema Marquee.

Vantagens	Desvantagens
<ul style="list-style-type: none"> <li>• Possibilidade de criação de <i>keywords</i>.</li> <li>• Criação de <i>keywords</i> e segmentos intuitiva.</li> </ul>	<ul style="list-style-type: none"> <li>• Impreciso na definição de segmentos (<i>timezones</i>).</li> <li>• Inexistência de reconhecimento de escrita de <i>keywords</i> para facilitar a procura no vídeo.</li> <li>• Espaço limitado para anotação.</li> </ul>

### 2.3.1.2 M4Note

Dos trabalhos relacionados, o sistema M4Note (MultiMedia Multimodal Annotation Tool) [1] é o sistema mais próximo da solução apresentada nesta dissertação. Similarmente, também possibilita a captura de um *stream* de vídeo e a anotação ao mesmo tempo que é capturado, por voz ou por tinta digital. As anotações são feitas através da interacção baseada em caneta digital em Tablet PC e podem ter duas abordagens diferentes: associação de metadados e enriquecimento do conteúdo.

A associação de meta-dados investiga o uso de modelos de meta-dados para construir uma estrutura semântica de forma a poder suportar operações como a indexação, procura e recuperação do conteúdo. Exige que o utilizador esteja ciente da estrutura semântica subjacente ao modelo do sistema, a fim de realizar anotações que estejam de acordo com a hierarquia de *tags*, o que pode ser uma tarefa difícil.

O enriquecimento do conteúdo faz-se via caneta digital para desenhar formas ou escrever texto e via microfone, para associar anotações de áudio. Contudo, não é possível executar operações como a recomendação e pesquisa por conteúdo e, por isso, este sistema converte o áudio e a tinta digital para texto, de forma a possibilitar tais operações.

Ambas as aproximações usam um modelo de metadados para permitir a reprodução de anotações sincronizadamente com o *media* capturado. Este modelo é implicitamente gerado e instanciado do modelo *MediaObject* [18]. *MediaObject* é um modelo baseado no standard MPEG-7, o que permite que as anotações sejam tratadas como objectos multimédia. Tal modelo torna possível a representação de descrições de baixo nível de forma automática, como o tamanho do vídeo e o tipo de compressão e a representação de descrições de alto nível, como a criação de *tags* personalizadas por utilizadores representando categorias, levando à criação de uma hierarquia de *tags*.

A ferramenta M4Note foi desenvolvida em Java, usando o Java Media Framework para controlar uma câmara web e o microfone incorporados no Tablet PC para capturar vídeo e áudio respectivamente. Além disso, também são capturados os metadados de baixo nível e descritos de acordo com o modelo *MediaObject*. No fim do processo de anotação, são gerados documentos XML como uma composição de referências de todos os *media* capturados: vídeo, áudio, imagens e texto. Quanto à interface, é composta por três elementos principais: o painel de visualização do vídeo capturado, a área de edição de anotações de tinta digital e a *timeline*, como ilustra a figura 2.6.



**Figura 2.6:** M4Note Interface [1]. No canto superior esquerdo encontra-se o painel de visualização do vídeo; no canto superior direito, encontra-se a área de edição de tinta electrónica para as *frames* seleccionados; em baixo encontra-se a *timeline* com os *thumbnails* das anotações.

O utilizador selecciona uma *frame* de vídeo para ser anotada clicando no painel de vídeo, que por sua vez suspende a captura. A *frame* seleccionada é convertida para uma imagem JPEG e apresentada na área de edição de anotações, onde o utilizador pode digitar com o teclado ou escrever à mão com a caneta directamente na imagem. Por baixo desta região, duas áreas mostram o texto que foi reconhecido via áudio (área à esquerda) ou via caneta digital (área à direita). O reconhecimento de voz usa a API do *Java Speech* (JSAPI) para comunicar com o software da IBM, ViaVoice™. O reconhecimento de escrita é feito utilizando uma técnica denominada por *Elastical Structural Matching Technique*, proposta por Chan e Yeung.

Um cenário de uso deste sistema é, por exemplo, numa experiência ao vivo como uma visita ao zoo, onde pode ser explorado por dois tipos de utilizador: um especialista, que usa o sistema para fazer explicações e comentários para os visitantes durante uma visita guiada e um visitante que queira registar a sua visita, muitas vezes relacionado com comentários gerais, notas e sentimentos pessoais acerca da mesma.

**Tabela 2.2:** Vantagens e desvantagens do sistema M4Note.

Vantagens	Desvantagens
<ul style="list-style-type: none"><li>• Possibilidade de criação de <i>tags</i>.</li><li>• Encapsulamento das anotações em XML.</li><li>• Conversão de áudio e tinta digital em texto.</li></ul>	<ul style="list-style-type: none"><li>• Inexistência de <i>motion tracking</i>.</li><li>• Anotação apenas de <i>frames</i> únicas em vez de segmentos.</li><li>• Limitação de espaço para a criação de anotações de tinta digital (apenas é possível criar anotações no topo da <i>frame</i>).</li></ul>

### 2.3.1.3 NoteLook

NoteLook [19] é um sistema cliente-servidor desenvolvido para suportar anotação multimédia com vídeo e tinta digital em reuniões/conferências. A aplicação é integrada numa sala de conferências equipada com três câmaras de vídeo e um projector. Os canais de vídeo contendo imagens da actividade dessa sala (*frames* do vídeo) e da apresentação (slides do projector) são transmitidas pelo servidor para os dispositivos de anotação dos clientes (Tablet PC's) durante a conferência/reunião e, consoante a preferência do utilizador, podem ser automaticamente incorporadas em páginas de notas (figura 2.7). Permite realizar anotações de tinta de formas livres nas imagens recebidas. Para o acesso e procura de notas e vídeos gravados, o NoteLook gera páginas Web com *links* das imagens e traços de tinta relacionados com o vídeo. Este sistema tem como predecessor o Dynamite, um sistema para captura e recuperação de notas de áudio e escritas.



**Figura 2.7:** Anotação numa imagem capturada e transmitida do servidor para este Tablet no sistema NoteLook [19].



**Tabela 2.3:** Vantagens e desvantagens do sistema NoteLook.

Vantagens	Desvantagens
<ul style="list-style-type: none"><li>• Captura automática de <i>slides</i> (quando os <i>slides</i> não contêm tipos dinâmicos como vídeo cliques).</li><li>• Geração de páginas Web com anotações associadas, o que permite a partilha das anotações entre os intervenientes da reunião e outros.</li></ul>	<ul style="list-style-type: none"><li>• Inexistência de anotação por voz.</li><li>• Dependente do bom funcionamento da rede.</li><li>• Dependente de uma sala equipada com tecnologia necessária.</li><li>• Elevado custo de manutenção.</li><li>• Anotação apenas de <i>frames</i> únicas em vez de segmentos</li></ul>

### Discussão

Dos trabalhos analisados, estes sistemas são os que mais se assemelham à solução proposta. No entanto, nenhum deles tem tantas funcionalidades como a solução proposta. Por exemplo, M4Note e NoteLook não têm a possibilidade de anotar segmentos de vídeo, que é um requisito importante na anotação de vídeo, pois cada *frame* é diferente de todas as outras e, como tal, anotar várias *frames* introduz mais significado ao vídeo. Além disso, M4Note introduz algumas limitações na interface, nomeadamente o facto da escrita de anotações de tinta digital estar restringida à área da *frame*, o que dificulta a sua visualização e diminui o espaço de criação. Em relação ao Marquee, tem uma forma de criação de *keywords/tags* criativa e fácil, que poderá ser usada como inspiração para o trabalho em curso.

### 2.3.2 Após a Captura do Vídeo

Estes sistemas assumem que existe um vídeo que já foi capturado anteriormente, onde os utilizadores irão associar anotações. São os sistemas mais comuns na actualidade e servem principalmente para enriquecer o conteúdo do vídeo e/ou prepará-lo para posterior procura e reutilização.

#### 2.3.2.1 LEAN

LEAN [20] é um protótipo que explora técnicas de interacção e visualização de navegação, segmentação, ligação e anotação de vídeos digitais, desenhado para o uso em Tablet PC sensíveis à pressão. Para visualização, este sistema permite a manipulação de um *stream* de vídeo, através de um conjunto de gestos efectuados com a caneta que permite suspender, reproduzir, parar ou ir até

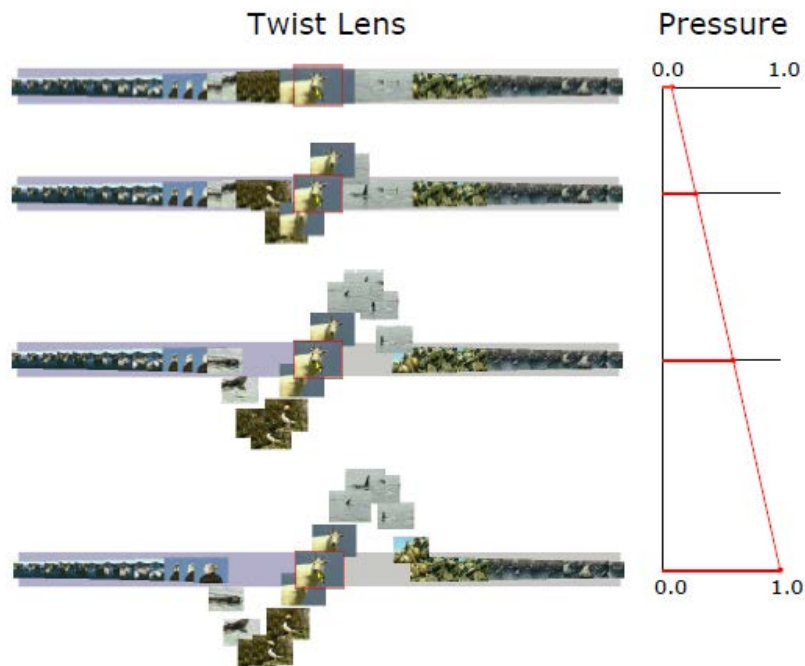
qualquer ponto no tempo do *stream* de vídeo. Estes gestos são feitos no conteúdo do vídeo para evitar desviar a atenção do utilizador na sua visualização. Além disso, o sistema permite aos utilizadores criar anotações e associá-las a uma *frame* ou a um segmento de vídeo. Existem dois controlos presentes neste sistema para visualização e manipulação de vídeo: *PVslider* e *TLslider*.

O *PVslider* é um controlo que permite aos utilizadores navegar num vídeo, arrastando a caneta horizontalmente para a direita ou para a esquerda para avançar ou recuar o conteúdo do vídeo respectivamente num dado intervalo, ou arrastando a caneta verticalmente para cima e para baixo para aumentar ou diminuir o tamanho desse intervalo. A velocidade com que o conteúdo do vídeo se move é constante até uma determinada posição. A partir dessa posição, a velocidade torna-se proporcional à distância entre essa posição e a posição inicial. O *TLslider* (*twist-lens*) é outro controlo que permite visualizar o *stream* de vídeo completo como uma sequência de *frames*, utilizando uma vista *fish-eye*, como ilustra a seguinte figura:



**Figura 2.8:** Controlo TLslider [20].

Para resolver o problema de *frames* estarem escondidas nesta vista, é alterado o aspecto da mesma criando uma forma em S na *frame* seleccionada e nas *frames* vizinhas. A amplitude da mesma varia de acordo com a pressão que é imposta pela caneta na *frame* seleccionada, como mostra a seguinte figura:



**Figura 2.9:** Controlo TLslider em forma de S [20]. Esta figura mostra, do topo para o fundo, como a amplitude do controlo varia com a pressão da caneta, que é ilustrada à direita.

As anotações criam-se escrevendo com a caneta numa área em branco da interface. Também existe a possibilidade de tornar uma anotação permanente, afixando-a em todo o intervalo de tempo do vídeo, pressionando a caneta sobre a anotação durante um intervalo de tempo definido.

**Tabela 2.4:** Vantagens e desvantagens do sistema LEAN.

Vantagens	Desvantagens
<ul style="list-style-type: none"> <li>• Técnicas de visualização e interação inovadoras na reprodução do vídeo e anotações associadas.</li> <li>• Interface simples e limpa.</li> <li>• Possibilidade de controlar o vídeo sem desviar a atenção do utilizador (<i>PVslider</i>).</li> <li>• Possibilidade de visualizar as <i>frames</i> do vídeo de forma criativa (<i>TLslider</i>).</li> <li>• Possibilidade de introduzir anotações permanentes.</li> </ul>	<ul style="list-style-type: none"> <li>• Não apropriado para vídeos de grande dimensão.</li> <li>• Dificuldade inicial de aprendizagem e interiorização de gestos.</li> <li>• Adaptação inicial ao controlo da pressão.</li> </ul>

### 2.3.2.2 Ambulant Annotator

Ambulant Annotator [5] é um sistema de anotação de vídeo multimodal para ser usado em Tablet PC em que, além de se poder anotar em forma de texto, *hyperlinks*, imagem, áudio ou tinta digital, também se podem animar as anotações inseridas. Este processo é manual e envolve o desenho de um *motion path* SVG que é aplicado a anotações individuais, como mostra a figura 2.10. O resultado da anotação e a representação do vídeo é exportado como um documento SMIL 2.0, sendo assim o formato de armazenamento para encapsular o vídeo e as suas anotações. O SMIL [21] tem várias vantagens como formato de armazenamento, como o facto de conter um rico conjunto de primitivas de *timing* de *media*, fornecendo uma arquitectura hipermédia para suportar *links* temporais para o conteúdo associado.

As anotações são feitas apenas para o enriquecimento do conteúdo do vídeo, sendo introduzidas de dois modos possíveis: de modo preemptivo e modo companheiro. Quando estão inseridas de forma preemptiva, as anotações suspendem a reprodução do vídeo quando são visualizadas. Quando estão inseridas em modo companheiro, as anotações são visualizadas paralelamente com a reprodução do vídeo.

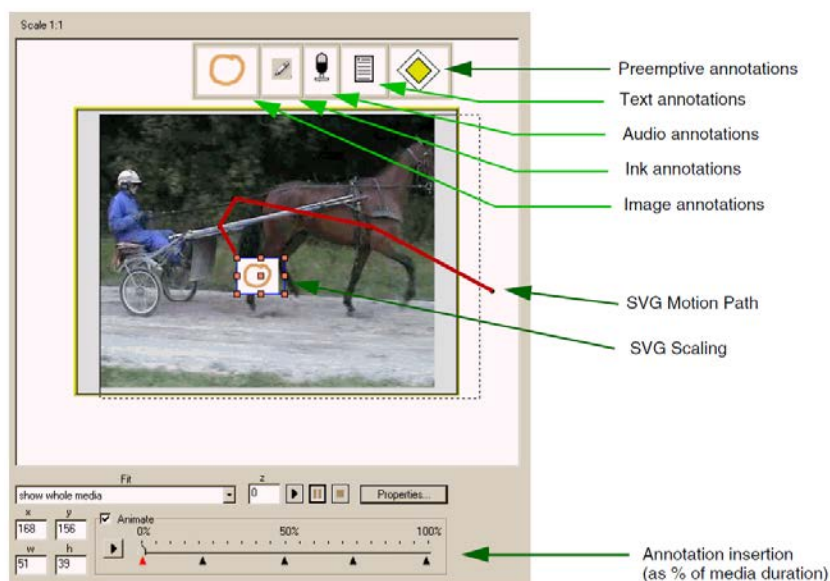


Figura 2.10: Interface do sistema Ambulant Annotator [5].

Tabela 2.5: Vantagens e desvantagens do sistema Ambulant Annotator.

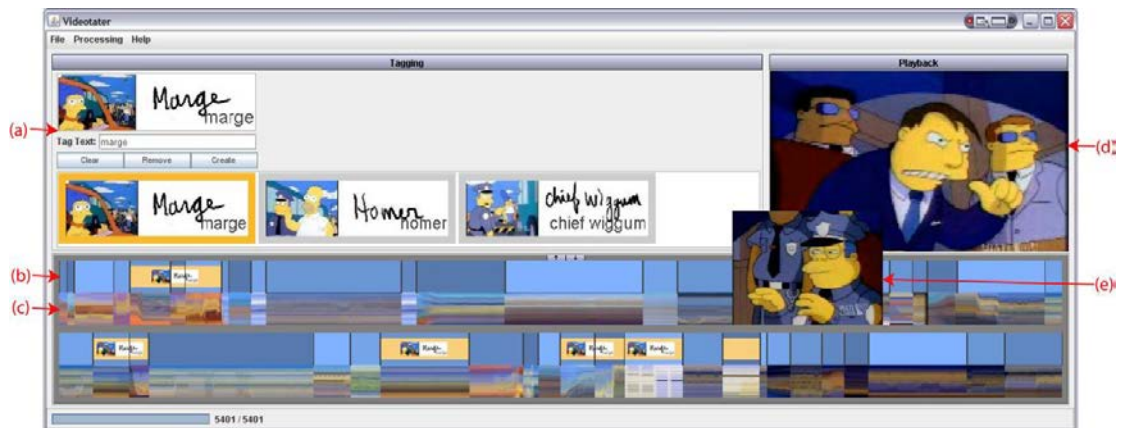
Vantagens	Desvantagens
<ul style="list-style-type: none"> <li>• Possibilidade de inserir anotações preemptivas.</li> <li>• Possibilidade de fazer <i>motion tracking</i> de objectos.</li> <li>• Uso de SMIL para encapsular anotações.</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Motion tracking</i> manual.</li> <li>• <i>Motion tracking</i> não apropriado para vídeos de média/grande dimensão.</li> <li>• Inexistência de uma visão geral das anotações presentes.</li> </ul>

### 2.3.2.3 Videotater

Videotater [22] é uma ferramenta experimental para Tablet PC, que suporta navegação, selecção, segmentação e *tagging* de vídeo. Foi desenhado para facilitar a preparação de vídeos de pequena dimensão (por exemplo episódios de séries de TV) para posterior reutilização, criando *tags* em segmentos definidos pelo utilizador ou em *frames* únicas.

Para inserir segmentos, o Videotater usa duas formas de visualização do conteúdo do vídeo. Uma forma de visualização é a *timeline* representada na figura 2.11, que mostra duas vistas diferentes do conteúdo do vídeo: a vista de segmentos na metade superior da *timeline*, onde os segmentos são delineados e apresentados e uma vista na metade inferior da *timeline*, onde é apresentada uma imagem em que cada coluna de pixels representa a média de cor da linha de uma *frame* no vídeo. Assim, torna-se mais fácil distinguir cada segmento pela diferença de cor entre segmentos. É uma *timeline* multi-linha, isto é, quando esta atinge a extremidade direita, continua na próxima linha em vez de

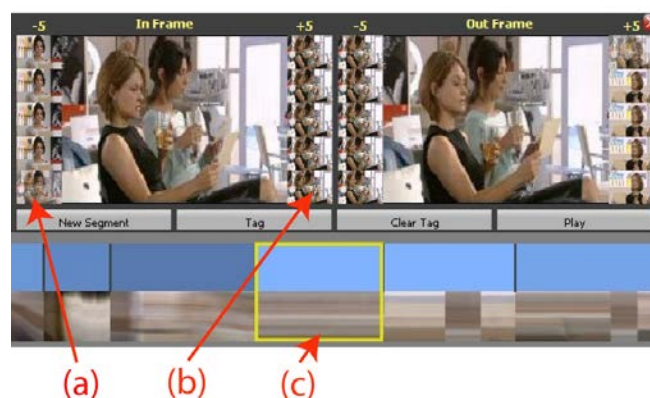
existir um *scroll* horizontal. Isto é benéfico para a visualização, pois existe uma visão geral do conteúdo do vídeo na interface, mas é apropriado apenas para vídeos de pequena dimensão.



**Figura 2.11:** Interface do sistema Videotater [22]: a) Criação de *tags*; b) Segmentos na *timeline*; c) imagem em banda que contém por cada coluna de píxeis a média da linha de uma *frame*; d) Visualização da reprodução do vídeo; e) *frame popup* da imagem em banda na *timeline*.

Existe um reconhecimento de gestos executados na *timeline*, que permite a rápida operação de juntar ou separar segmentos. Ao desenhar uma linha vertical sob um segmento, o mesmo é separado no instante indicado pela média da coordenada *x* dessa linha. Ao desenhar uma linha entre dois segmentos, é feita a junção de ambos. O modo de selecção e o modo de *tagging* de segmentos é alterado consoante a pressão da caneta seja de baixa ou alta intensidade respectivamente.

A outra forma de visualização é a visualização polifocal (*Polyfocal Visualization*) e é apresentada quando um segmento na *timeline* é seleccionado. Esta visualização está representada na figura 2.12:



**Figura 2.12:** a) 5 frames anteriores à *in frame*; b) 5 frames posteriores à *in frame*; c) segmento seleccionado [22].

Esta vista contém do lado esquerdo a *frame* inicial (*in point*) e do lado direito a *frame* final (*out point*) do segmento seleccionado na *timeline* (c). Em cada uma destas *frames*, é também apresentado um conjunto de 5 *frames* antes (a) e depois da *frame* corrente (b), representando o seu

contexto. A partir desta visualização, pode-se criar ou editar um segmento e associar uma *tag* ao mesmo.

**Tabela 2.6:** Vantagens e desvantagens do sistema Videotater.

Vantagens	Desvantagens
<ul style="list-style-type: none"> <li>• Identificação e gestão facilitada de segmentos do vídeo para serem anotados.</li> <li>• Existência de <i>timeline</i> detalhada para uma visão geral do conteúdo do vídeo.</li> <li>• Uso da pressão da caneta para alternar entre modos de interacção.</li> </ul>	<ul style="list-style-type: none"> <li>• Não apropriado para vídeos de média/grande dimensão.</li> <li>• Não suporta anotações de áudio.</li> <li>• Não suporta <i>motion tracking</i>.</li> </ul>

## Discussão

Em relação a estes sistemas, o contexto e propósito de cada um é diferente. Enquanto que o LEAN se destina a explorar e apresentar técnicas de visualização e manipulação de vídeo, o Ambulant Annotator visa a anotação de vídeo para enriquecer o seu conteúdo e o Videotater para preparação para a procura e reutilização de vídeo. Nestes sistemas, existem técnicas de manipulação e visualização interessantes. Em relação ao LEAN, os gestos executados com a caneta para controlo de vídeo são vantajosos relativamente aos controlos usuais, que normalmente estão fora do conteúdo do vídeo, e desviam a atenção do utilizador em relação ao mesmo. Em relação ao Ambulant Annotator, existe a noção de anotação preemptiva, que pode ser importante na visualização de anotações, se a mesma se destinar a uma *frame* em particular ou se a mesma requerer uma atenção especial para a sua leitura ou compreensão. No Videotater, o destaque está na metade inferior da *timeline*, que contém a divisão visual automática de segmentos, que pode ser útil na sua identificação para anotação.

## 2.4 Interacção com Caneta Digital e Multi-Toque

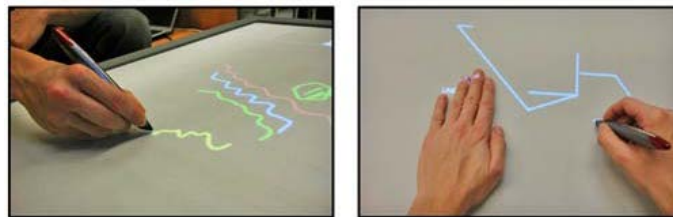
A interacção directa com os ecrãs de dispositivos está-se a tornar rapidamente num dos principais meios pelo qual as pessoas experienciam a computação. Este facto levou a um interesse renovado em dispositivos que suportem multi-toque e caneta digital. Ao existir input das duas mãos, os utilizadores conseguem aumentar a largura de banda do input e adicionar interacções ricas e naturais às suas aplicações [23]. Esta interacção bimanual tem sido investigada em vários projectos como os descritos a seguir.

### 2.4.1 Benefits of Bimanual Pen and Direct-Touch Interaction on Horizontal Interfaces

Este artigo [23] defende o uso do input da caneta digital e toque humano feito com as duas mãos, descrevendo uma experiência em que os utilizadores testam os inputs caneta/toque, toque/toque e caneta/caneta. Além disso, apresenta princípios de design deste tipo de input e um protótipo que dá uso a esses princípios.

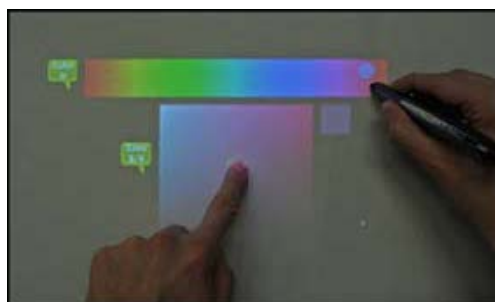
Quando se usa a caneta como input, distingue-se tipicamente entre dois modos: o modo de tinta ou o modo de comando, isto é, referindo-se ao uso da caneta para escrever/desenhar ou para executar comandos respectivamente. Quando se usa o toque, este é interpretado normalmente como uma interacção baseada no apontador, como se fosse uma interacção com o rato.

No protótipo elaborado, os autores propuseram o uso da caneta com a “mão dominante” (DH – *dominant hand*) para input preciso e o toque com a “mão não dominante” (NDH – *non-dominant hand*) para comandos. Por exemplo, para desenhar é usada a caneta, pois tem mais precisão e provoca menos oclusão no ecrã e o toque é usado para comandos, como a gestão da área de desenho através do *zoom* e do *panning* (arrastamento) dessa área. A figura 2.13 demonstra que se o utilizador toca na superfície enquanto a caneta está no modo de desenho, o modo de desenho passa de formas livres para linhas rectas, demonstrando que o toque tem como objectivo a execução de comandos, neste caso alternância de modos da caneta.



**Figura 2.13:** Desenho em formas livres (esquerda); Desenho com linhas rectas quando existe toque (direita) [23].

Outro exemplo é na selecção de cores no sistema de cores HSV (*Hue Saturation Value*), em que o utilizador escolhe o H com a caneta de forma precisa e o S e o V com o toque, por isso estes componentes são escolhidos simultaneamente existindo assim mais largura de banda no input, como ilustra a figura 2.14.



**Figura 2.14:** Escolha de cor bimanual [23].



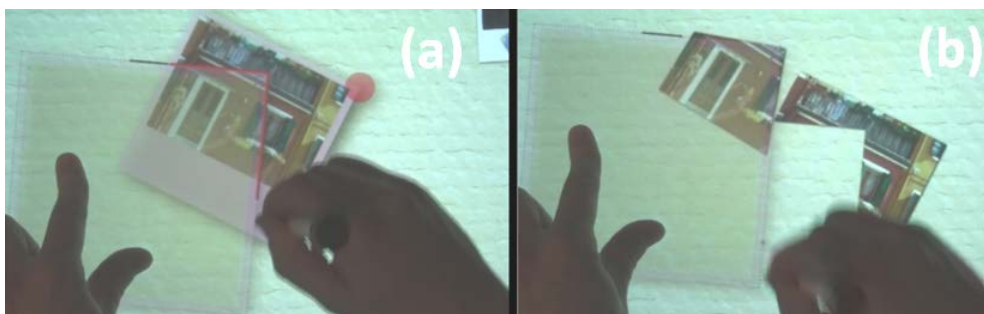
Realizou-se uma experiência para saber quais as diferenças entre as combinações de pares de input existentes entre caneta e toque. Para isso, 12 utilizadores, com idades compreendidas entre 20 e 50, tinham como tarefa traçar um caminho num labirinto desde o ponto de partida até ao ponto de chegada sem tocar nas paredes. Cada participante testou as 3 técnicas possíveis: caneta/caneta, caneta/toque e toque/toque. O labirinto estava desenhado para que, para além do desenho do traço, os utilizadores tinham de fazer *zoom* e *panning* ao labirinto com o toque para concluírem a tarefa com êxito.

Concluiu-se que a técnica caneta/toque foi a mais bem sucedida em termos de tempo a completar a tarefa, erros cometidos e preferência do utilizador. Os utilizadores demoraram menos tempo a completar a tarefa, cometeram menos erros, isto é, o traço atingiu menos vezes as paredes, e foi a técnica preferida em relação às restantes.

#### 2.4.2 Manual Deskterity

Manual Deskterity [24] executa-se na plataforma Microsoft Surface e é uma aplicação que explora a caneta digital e toque para servir como um estúdio virtual de um *designer*, abstraindo uma secretária e um bloco de notas pessoais.

Nesta aplicação, a caneta escreve, o toque manipula e a combinação da caneta e toque dá origem a novas ferramentas. A combinação da caneta e toque traduz-se na transição automática do modo da caneta de escrita para o modo de comandos gestuais, quando é seleccionado um objecto com dois dedos. Por exemplo, para cortar uma imagem, ilustrada pela figura 2.15, o utilizador selecciona um objecto com os dois dedos que pode ser usado como régua para o corte (a). Assim, ao desenhar com a caneta ao longo dessa régua e por cima da imagem alvo, a imagem alvo vai ser cortada em duas imagens (b) sendo o corte efectuado no traço da caneta.



**Figura 2.15:** Corte ao longo da régua (a); A imagem é cortada em dois pelo corte desenhado (b) [24].

Existem outras tarefas nesta aplicação que exploram este método de interacção, como a navegação de páginas de álbuns, *brushing* e cópia de imagens. Os utilizadores consideraram que esta aproximação da caneta e toque é apelativa e que o padrão geral da aplicação (seleccionar com o toque + gestos com a caneta) é comum à maior parte dos nossos gestos.



## Discussão

Estes projectos demonstraram os benefícios da interacção bimanual para uma variedade de tarefas. Quando os designers escolhem esta interacção como input, é preciso saber que mão irá controlar que dispositivo de entrada. Segundo a literatura [23], a utilização de cada dispositivo individual (como o rato, caneta ou mesa multi-toque) tem vantagens e desvantagens, por isso, é necessário considerar estes *trade-offs* tendo em conta o tipo de aplicação a que se destinam. Por exemplo, em superfícies multi-toque, existe uma grande vantagem de que nenhum input intermediário é necessário e, por isso, este tipo de interacção directo é chamado de natural quando comparado com uma interacção com o rato ou caneta. Adicionalmente, ao existirem múltiplos pontos de contacto, estes dispositivos fornecem um input mais complexo. Por outro lado, uma caneta digital proporciona uma maior precisão, rigor, mas tipicamente só existe um ponto de contacto. Existem vantagens e desvantagens na utilização da caneta digital ou do toque como input, como ilustra a tabela 2.7:

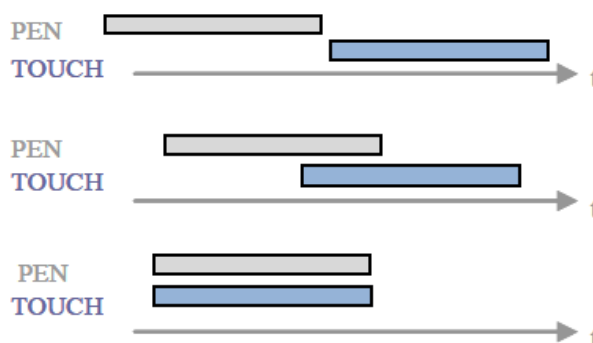
**Tabela 2.7:** Vantagens e desvantagens no uso da caneta ou de toque como input [23].

	<b>Vantagens</b>	<b>Desvantagens</b>
<b>Input de Caneta</b>	<ul style="list-style-type: none"><li>• Input menos acidental do que o toque.</li><li>• Ponto de toque mais preciso/ Resolução elevada.</li><li>• Ferramenta familiar que influencia a experiência dos utilizadores.</li><li>• Menos oclusão de alvos do que o toque.</li></ul>	<ul style="list-style-type: none"><li>• Apenas um ponto de input.</li><li>• Dispositivo separado.</li><li>• Menos oclusão enfatiza paralaxe.</li></ul>
<b>Input de Toque</b>	<ul style="list-style-type: none"><li>• Múltiplos pontos de input/ alta largura de banda de inputs.</li><li>• Usado com pouca atenção.</li><li>• "Natural".</li><li>• Nenhum input extra para interagir.</li></ul>	<ul style="list-style-type: none"><li>• Oclusão pelas mãos ou dedos.</li><li>• Baixa precisão no ponto do toque.</li></ul>

A combinação entre estes dois inputs é que pode dar resultado a uma série de novas formas de interacção que, além de permitir usar as duas mãos para input executando tarefas simultâneas, dá origem a mais fluidez e naturalidade no uso de interfaces aproximando-se ao “papel electrónico”. No fundo, a vantagem de se usar esta combinação é o facto de dar a possibilidade de interagir mais

precisamente com a caneta, por exemplo escrever, e menos precisamente e mais informalmente com o toque, como por exemplo manipular a interface acedendo a botões ou menus.

Existem muitas formas de completar uma tarefa usando estes dois inputs que, na maioria, se reflectem no contexto onde são usados, como representa a figura 2.16. O facto de se usar primeiro a caneta seguido do toque ou vice-versa, ou usando os dois em simultâneo/em sobreposição para completar uma tarefa dá origem a contextos diferentes, que podem ser usados para diferenciar certas acções.



**Figura 2.16:** Três sequências possíveis (sequencial, sobreposto e simultâneo) que causam comandos diferentes [23].

Claro que existem tarefas onde não é necessário usar esta combinação, nomeadamente tarefas simples em que só um input é suficiente. Por isso, é preciso ter em conta esta complexidade por vezes desnecessária. Por outro lado, esta interação pode introduzir um problema chamado de “rejeição da palma da mão”, em que o utilizador descansa a sua mão no ecrã enquanto escreve, levando potencialmente a operações não intencionais. Esse problema resolve-se tratando os contactos de grande área como acidentais e, por isso, sem efeito.

Em suma, esta combinação entre caneta e toque é útil, proporcionando mais fluidez, familiaridade com o papel e rapidez na execução de tarefas, dando origem um vocabulário de input expandido. A caneta digital e o toque têm vantagens e desvantagens próprias, no entanto eles complementam-se e, por isso é usada a combinação entre eles, dando a possibilidade de interagir com as mais variadas interfaces num Tablet PC.

No sistema apresentado nesta dissertação, esta combinação vai ser usada, no sentido em que a caneta serve para escrever anotações de tinta digital (escrita à mão ou desenhos livres) e o toque para interagir com os *widgets* da interface gráfica. Adicionalmente, pode ser usada a pressão e a execução de gestos com a caneta para executar comandos, bem como a exploração de combinações dos dois inputs para realizar tarefas, desde que não introduza complexidade na sua execução.

### **3. Captura e Anotação de Vídeo**

O estudo dos trabalhos relacionados permitiu investigar a área de contribuição, assim como o contexto e motivação onde a dissertação se insere. Novas ideias e conhecimentos foram obtidos, os quais se revelaram necessários para um melhor desenvolvimento da solução proposta. Neste capítulo, apresentam-se os conceitos e as tecnologias utilizadas no desenvolvimento do sistema proposto, assim como os requisitos impostos. A secção inicial foca o contexto do problema e são apresentados os principais requisitos a cumprir. De seguida, são apresentados os mecanismos que visam satisfazer os requisitos anteriormente especificados, incluindo a descrição da interface, assim como as funcionalidades oferecidas pelo sistema proposto.

#### **3.1 Conceito**

O projecto TKB (*Transmedia Knowledge Base*), que define o âmbito desta dissertação, tem como um dos principais objectivos o desenvolvimento de uma ferramenta orientada ao processo de criação e arquivos individuais dos coreógrafos, intitulada “Creation-Tool”. Esta ferramenta é a solução proposta nesta dissertação. Trata-se de um anotador de vídeo que suporta a captura e anotação multimodal de vídeo em tempo real, aplicado à dança contemporânea como uma ferramenta de criação.

Como a dança contemporânea é um domínio onde a captura e anotação de vídeo é relevante a fim de enriquecer e melhorar os processos criativos, esta ferramenta foi concebida e desenhada para assistir o trabalho dos coreógrafos durante os seus ensaios, funcionando como um bloco de notas digital que contém anotações pessoais relativas aos mesmos. Com base nesta informação, na forma de notas, associada com os vídeos dos ensaios, é possível construir um vasto arquivo de conhecimento e aprendizagem, em que os artistas possam rever o seu trabalho e, assim, aperfeiçoar tanto a sua performance, como o espectáculo em si. A utilização alvo da solução é essencialmente centrada durante um evento ao vivo, contudo também pode ser útil depois do mesmo, actuando como uma ferramenta de criação, mas sobretudo como uma análise do trabalho anteriormente realizado.

O Tablet PC é a tecnologia escolhida para executar esta ferramenta, uma vez que emprega a combinação da interacção baseada na caneta digital com o toque, que se aproxima da interacção humana com o papel e caneta. Por ser natural, esta interacção pode ser aplicada em qualquer ambiente, sem interferir com processos criativos existentes e metodologias de trabalho. Quando é utilizado de forma móvel, o Tablet PC é segurado com uma mão, que corresponde à mão não dominante. A mão não dominante agarra uma extremidade lateral para mudar de modo na ferramenta e a mão dominante sustém a caneta digital para escrever notas.

Esta ferramenta diferencia-se de trabalhos anteriores, exactamente por permitir a captura de vídeo simultaneamente com a criação de anotações via caneta digital, toque humano ou voz. Este conjunto de modalidades de input, juntamente com as anotações em tempo real e métodos de *motion tracking*, não está presente nos trabalhos relacionados, por isso esta solução contribui substancialmente para melhorar a análise e anotação do conteúdo do vídeo, particularmente quando aplicado à dança contemporânea.

Todo o processo de desenho, concepção e avaliação da ferramenta foi e será acompanhado por dois coreógrafos de dança contemporânea: Stephan Jürgens<sup>6</sup> e Rui Horta<sup>7</sup>. Adicionalmente, a *designer* que contribuiu mais para a interface de utilizador é também uma bailarina e coreógrafa.

## 3.2 Requisitos

Os requisitos impostos consistiram em condições definidas pelos utilizadores e equipa de projecto, com o objectivo de tornar a solução proposta mais desejada e adequada. Esta definição foi estabelecida no início, mas ao longo do desenvolvimento, foi constantemente actualizada conforme as circunstâncias encontradas. Neste capítulo, são referidos os requisitos de desenho, ou seja, as características e regras que a interface deve atender e os requisitos funcionais, isto é, as operações exigidas que a solução deve implementar.

### 3.2.1 Requisitos de Desenho

O desenho da interface tem sido acompanhado por dois coreógrafos com diferentes necessidades e processos de criação. Os coreógrafos, os bailarinos e os especialistas em tecnologia aplicada à dança foram envolvidos desde o início, para a obtenção mais apropriada dos requisitos da ferramenta.

A usabilidade, o desenho *user-friendly* e a capacidade de categorizar a informação numa estrutura coerente são os principais desafios para o processo de criação da interface. A estrutura de

---

<sup>6</sup> Biografia de Stephan Jürgens: [http://www.estc.ipl.pt/teatro/paginas\\_profes/stephan\\_jurgens.html](http://www.estc.ipl.pt/teatro/paginas_profes/stephan_jurgens.html)

<sup>7</sup> Biografia de Rui Horta: [http://www.oespacodotempo.pt/pt/esp\\_tem.php?idpan=rui\\_bio](http://www.oespacodotempo.pt/pt/esp_tem.php?idpan=rui_bio)

navegação deve ser desenvolvida baseada em dois níveis de interacção pessoa-máquina: Hardware (interface física - Tablet PC) vs. Software (interface virtual - conteúdos). É preciso ter em conta que o tamanho do dispositivo incentiva à interacção bimanual, que promove mudanças de modo entre a caneta digital e o toque e acelera a navegação nos conteúdos da interface.

Os requisitos principais para o desenho da interface prendem-se com a criação de uma interface pensada de raiz para ser executada em Tablet PCs, com o menor ruído gráfico possível usando cores neutras (através de uma escala de cinzentos, que evita a fadiga ocular), seguindo o padrão das interfaces de aplicações de vídeo. Além disso, os elementos gráficos da interface devem estar separados de todo o conteúdo visual restante. Os tamanhos e espaços entre estes elementos estão directamente relacionados com as especificações de ecrãs de toque, de forma a terem uma melhor performance [25].

Quanto aos elementos da interface, devem estar agrupados em várias áreas distintas para uma melhor organização. Deve existir uma área onde são criadas notas no topo ou em redor do vídeo, uma área contendo um menu de sistema, uma área que diz respeito à gestão do projecto, uma área de fácil acesso para escolher as ferramentas de anotação a utilizar e, finalmente, uma forma de visualização e navegação do conteúdo do vídeo e respectivas anotações.

### **3.2.2 Requisitos Funcionais**

O principal requisito funcional imposto prende-se com a capacidade de captura de vídeo e anotação em tempo real.

Capturar um vídeo é uma operação que exige alguma complexidade. A gravação de imagens digitais (*frames*) em movimento, que resultam numa animação sequencial é uma tarefa computacionalmente pesada, pois envolve a manipulação e acesso a imagens, em que cada uma representa uma matriz de píxeis que são analisados sucessivamente. Além deste factor, a tarefa de construir um vídeo, com base em imagens, é executada numa sequência de três passos principais. Primeiro, é necessário captar as imagens provenientes da fonte de vídeo. De seguida, essas imagens são armazenadas e adicionadas ao conjunto de imagens já anteriormente recebidas. Finalmente, é adicionado, caso tenha sido capturado, áudio sincronizado ao objecto de vídeo, o qual é gravado num sistema de ficheiros. Este problema deve ser resolvido de forma a que não interfira com a tarefa de anotação em tempo real e, ao mesmo tempo, produza um vídeo com uma *framerate* e qualidade aceitáveis. Estes requisitos aplicam-se igualmente à captura de áudio.

A anotação de vídeo em tempo real é uma funcionalidade que, para originar um bom resultado, deve responder a vários desafios. O principal desafio é fazer com que, no final da captura, o utilizador tenha anotado cada momento chave do vídeo, que tenha sido um potencial alvo para associação de anotações. Na anotação em tempo real, existem restrições que não se verificam na

anotação de um vídeo capturado, onde se tem uma noção e controlo total do seu conteúdo. Em tempo real, o utilizador somente tem conhecimento dos momentos passados e presentes do vídeo, além de estar a visualizar o vídeo pela primeira vez, o que torna a tarefa de anotação mais penosa. Assim, torna-se essencial criar mecanismos para a percepção e posterior anotação de todos os momentos relevantes sujeitos a anotações. Além disso, é necessário ter métodos de inserção de anotações rápidos, simples e eficazes que adicionem a informação pretendida e que, ao mesmo tempo, não desviem a atenção do utilizador do que se passa no vídeo enquanto o processo de anotação decorre. Estas são as soluções que resolvem o desafio, tornando-se requisitos cruciais para que esta funcionalidade seja eficiente.

Quanto aos tipos de anotação, é importante que sejam naturais, de fácil aprendizagem, para que se tornem transparentes ao processo criativo. Adicionalmente é fundamental que estas anotações sejam armazenadas separadamente do vídeo, ou seja num formato independente, com o objectivo de serem reutilizáveis noutros vídeos. Desta forma, as anotações são uma camada que se sobrepõe ao vídeo, em vez de serem embutidas dentro do próprio vídeo, o que danificaria o seu conteúdo original.

O movimento, por exemplo do corpo humano, é um elemento presente e essencial na dança e as anotações de vídeo gráficas, particularmente as que são desenhadas usando a caneta digital, podem exigir métodos de um acompanhamento desse movimento (*motion tracking*). Assim, quando for requisitado, este tipo de anotações deve seguir automaticamente os elementos de vídeo a que estão associados, de forma a manterem o seu contexto no conteúdo do vídeo.

Outro requisito importante para esta solução é a mobilidade. Se o utilizador é um coreógrafo de dança, ele pode querer mover-se livremente, por exemplo, durante um ensaio e registar, ainda assim, informação sobre ele. Consequentemente, devem existir formas de acesso remotas à ferramenta, para que o coreógrafo não tenha de estar sempre em contacto com o Tablet PC durante a captura do evento.

Finalmente, será desejável se a tecnologia de software onde for desenvolvida esta solução tiver suporte para multi-plataforma, caso, no futuro, seja necessário criar outras versões para plataformas diferentes. Este requisito torna-se importante quando a solução é aplicada à dança, dado que os artistas envolvidos possuem, na sua maioria, computadores com o sistema operativo Mac OSX.

### **3.3 Solução**

Durante o processo de criação do coreógrafo, uma vasta quantidade de documentos, como anotações, vídeos e referências são gerados, compilados e organizados num suporte. Esta ferramenta compromete-se a ser esse suporte. Com base nos requisitos especificados, pretende-se desenvolver uma interface gráfica direccionada para Tablet PC que possibilite a captura e anotação de vídeo em tempo real, oferecendo uma experiência fluida, acessível, cooperante, criativa e enriquecedora. Neste

capítulo, são apresentadas as soluções práticas em resposta aos requisitos de desenho e funcionais, incluindo as tecnologias utilizadas, descrição da interface e respectivas funcionalidades.

### 3.3.1 Tecnologias

As tecnologias de software usadas para o desenvolvimento da solução proposta foram as plataformas Qt GUI, para o desenho e implementação da interface, e openFrameworks para o desenvolvimento das funcionalidades, como a captura, armazenamento e reprodução do vídeo e anotações.

A plataforma Qt GUI permite criar interfaces em C++ para multi-plataforma, incluindo sistemas desktop, como *Windows*, *Linux* ou *Mac*, ou sistemas embebidos, como *Symbian*. Contém um SDK, que é constituído por várias ferramentas de desenvolvimento, entre as quais o Qt Designer. O Qt Designer permite desenhar graficamente a interface num *form*, acrescentando-lhe *widgets*. Um *widget* é um elemento gráfico da interface, como um botão, *check box* ou uma caixa de texto. No final deste processo, gera um ficheiro C++ com o código resultante.

A plataforma openFrameworks é uma biblioteca *open-source* multi-plataforma em C++, concebida para programação criativa de aplicações interactivas. Recorre a bibliotecas multimédia poderosas que permitem aceder e manipular texto, imagem ou vídeo de uma forma simples e eficiente. Como é *open-source*, tem estado em constante desenvolvimento, existindo cada vez mais *addons* criados que complementam a plataforma, para resolver problemas que surgem durante o desenvolvimento de projectos.

Contudo, estas plataformas são independentes, por isso, houve uma necessidade de integrá-las num só *framework*, para que funcionassem em conjunto. Assim, inicialmente foi finalizada esta tarefa, que já tinha sido iniciada pelo doutorando Diogo Cabral. A partir deste contributo que acreditamos ser pioneiro, já que não existem vestígios de guias oficiais para tal, foi realizado um manual<sup>8</sup> que se tornou público contendo os passos necessários para a integração destas duas plataformas nos IDEs Visual Studio 2008 e 2010. Com esta integração, é possível construir aplicações multimédia que, além de suportarem a programação criativa com o openFrameworks, disponibilizando o acesso e manipulação de documentos multimédia de uma forma simples e intuitiva, proporcionam uma interface gráfica profissional e normalizada para o utilizador através do Qt GUI. Isto vem preencher lacunas das duas plataformas, que eram por um lado o acesso limitado a documentos multimédia na plataforma Qt GUI e, por outro, o suporte para a elaboração simplificada de interfaces gráficas profissionais na plataforma openFrameworks. Estes são os requisitos da solução, portanto, este *framework* Qt+openFrameworks, usando a biblioteca openCV para processamento de vídeo foi a tecnologia que se julgou mais adequada para implementar a proposta de funcionalidades da solução.

A tecnologia de hardware, onde a solução foi desenvolvida e testada é um Tablet PC HP TouchSmart tm2-2150ep. Contém um processador Intel Core i3-380UM 1.33 GHz, memória de 4 GB

---

<sup>8</sup> [http://img.di.fct.unl.pt/diogocabral/QT+OF\\_ProgrammersGuideC++.pdf](http://img.di.fct.unl.pt/diogocabral/QT+OF_ProgrammersGuideC++.pdf)

DDR3, disco rígido SATA de 500 GB, e uma placa gráfica ATI Mobility Radeon HD 5450 com 512 MB DDR3 reservados. O ecrã é táctil conversível de 30,7 cm (12,1 polegadas) e tem uma resolução máxima de 1280x800, que, de resto, foi a resolução escolhida para a ferramenta. Possui ainda um teclado, 3 portas USB 3.0, uma câmara web HP TrueVision com microfone digital integrado e peso de 1,89 kg. Além do Tablet PC, o dispositivo iPod Touch 3G serviu de controlo remoto da ferramenta e o microfone *Bluetooth* Nokia BH-05 como entrada de áudio para captura de anotações de voz e som ambiente do vídeo.



**Figura 3.1:** Diferentes tecnologias de input da ferramenta.

Para o armazenamento das anotações, bem como configurações de projecto, foi criada uma base de dados em formato XML.

### **3.3.2 Interface**

De acordo com os requisitos, foram desenvolvidos dois tipos de interface: (1) a interface principal executada no Tablet PC, que representa a interface da solução proposta e (2) a interface do dispositivo móvel presente no controlo remoto. Cada interface é apresentada separadamente nesta secção.

#### **3.3.2.1 Interface do Tablet PC**

A interface principal corresponde à interface da ferramenta presente no Tablet PC. Para não provocar fadiga ocular no utilizador, já que o mesmo poderá ocupar bastante tempo a utilizar a ferramenta pelo facto de trabalhar com vídeo, foram criadas duas versões da interface com cores neutras em escala de cinza. Inicialmente, foi desenvolvida a versão *dark*, que usa uma escala de cor escura que varia



aproximadamente entre os valores RGB(0,0,0) até RGB(150,150,150). Quando foi testada esta versão, verificou-se que, por vezes, provocava alguma dificuldade em perceber os elementos gráficos presentes, devido ao reflexo da luz no ecrã. Assim, foi desenhada outra versão, chamada *light*, com uma escala de cor mais clara que varia sensivelmente entre os valores RGB(150,150,150) até RGB(255,255,255). Esta versão resolveu as dificuldades verificadas, tornando-se a versão própria para uso em ambientes mais claros. A versão *dark* é ideal para uso em ambientes fechados e mais escuros, onde a luz não tenha uma forte intensidade que diminua a visibilidade da interface no ecrã.

A interface principal, ilustrada na figura 3.2, encontra-se organizada e distribuída em cinco secções distintas (do topo para o fundo):

- **Menu de Sistema** (*system menu*): fixado no canto superior esquerdo, com as funcionalidades de sistema padrão (*open, save, edit, help*);
- **Menu de Projecto** (*project menu*): localizado no canto superior direito, gere todo o conteúdo e ficheiros gerados durante o processo de anotação dentro de cada projecto.
- **Barra de Ferramentas** (*toolbar*): situada no lado esquerdo, contém as ferramentas de anotação e activação e modos de captura de vídeo (de cima para baixo: *capture modes, marks, draw, text, audio, link, motion tracking, cameras*). A *toolbar*, sendo o menu mais frequentemente acedido, encontra-se do lado esquerdo, para permitir que utilizadores destros possam abordá-lo facilmente com a mão esquerda, enquanto a mão direita segura a caneta digital.
- **Barra de Ferramentas Auxiliares** (*toolsets*): consistem em propriedades e transformações que se podem aplicar às anotações criadas (de cima para baixo: *selection, color picker, eraser, line thickness, type font, group/ungroup, copy, delete*).
- **Timeline**: posicionada por baixo da visualização de vídeo, é usada para a navegação do vídeo e anotações. A *timeline* pode ser arrastada para cima ou para baixo, evitando a oclusão de parte da *toolbar*, conteúdo do vídeo e anotações.

Ao centro, encontra-se a visualização do vídeo com uma largura de 640 pixéis e altura de 480 pixéis. Em redor da visualização, não existem elementos gráficos da interface, de modo a que o utilizador tenha espaço para criar as suas anotações, já que elas podem ser criadas dentro ou fora do conteúdo do vídeo. Para melhorar o *feedback* ao utilizador usando a interacção baseada na caneta digital e toque, todos os elementos gráficos da interface têm uma resposta imediata no momento em que o utilizador os pressiona.

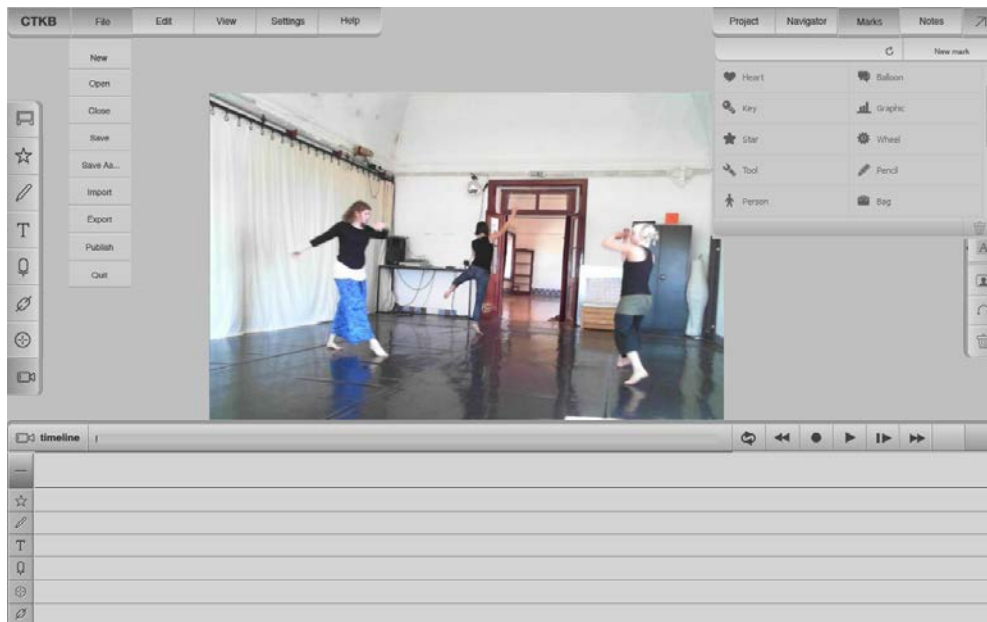


Figura 3.2: Interface versão *light* da ferramenta no Tablet PC.

### 3.3.2.2 Interface do Dispositivo Móvel

Para responder ao requisito de mobilidade, ou seja, de oferecer mais liberdade de movimentos e gestos essencial em ambientes de espectáculo de dança, algumas operações podem ser controladas remotamente através de um dispositivo móvel via OSC (*Open Sound Protocol*). OSC, originalmente planeado para partilhar dados de performance musical, é um protocolo de comunicação entre computadores, sintetizadores de som ou dispositivos multimédia, que permite trocar mensagens simples através de uma rede por UDP/IP [26].

Foi desenhada uma interface na aplicação *desktop* multi-plataforma TouchOSC Editor 1.5.2<sup>9</sup>, que permite criar e enviar interfaces para dispositivos móveis do tipo Android ou iOS. Já a aplicação móvel TouchOSC, residente no dispositivo móvel correspondente, recebe, guarda e posteriormente executa no dispositivo as interfaces criadas.

O dispositivo móvel torna-se assim um controlo remoto, que tem como função executar até duas operações: iniciar ou terminar uma captura de vídeo (*Record*) e iniciar ou terminar uma captura de áudio (*Audio Note*), que corresponde a uma anotação de voz. Assim, este controlo usado juntamente com um microfone *Bluetooth*, possibilita ao utilizador, remotamente e sem fios, iniciar a captura de vídeo, criar anotações por voz e, no fim, terminar a sessão. Cada operação tem um botão correspondente que, quando é pressionado, a inicia. Quando o botão é pressionado pela segunda vez, termina a mesma operação.

Este tipo de controlo é benéfico quando o coreógrafo não se encontra em contacto com o Tablet PC e, ainda assim, deseja iniciar ou continuar o seu processo de anotação.

<sup>9</sup> <http://hexler.net/software/touchosc>

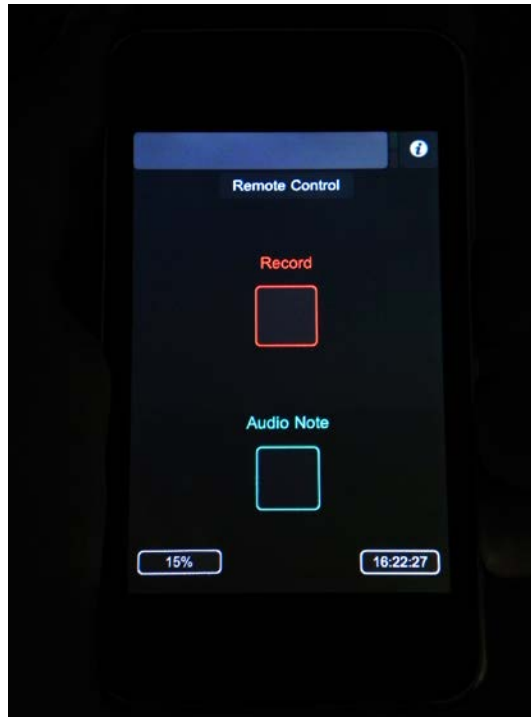


Figura 3.3: Interface do dispositivo móvel executada num iPod Touch 3G.

### 3.3.3 Funcionalidades

As funcionalidades têm por objectivo satisfazer os requisitos impostos da ferramenta. São as respostas na forma de soluções práticas para resolver os problemas descritos anteriormente. Neste capítulo, são descritas as funcionalidades principais da ferramenta, desde a captura de vídeo e os tipos e modos de anotação existentes até à gestão de cada projecto criado.

#### 3.3.3.1 Captura de Vídeo

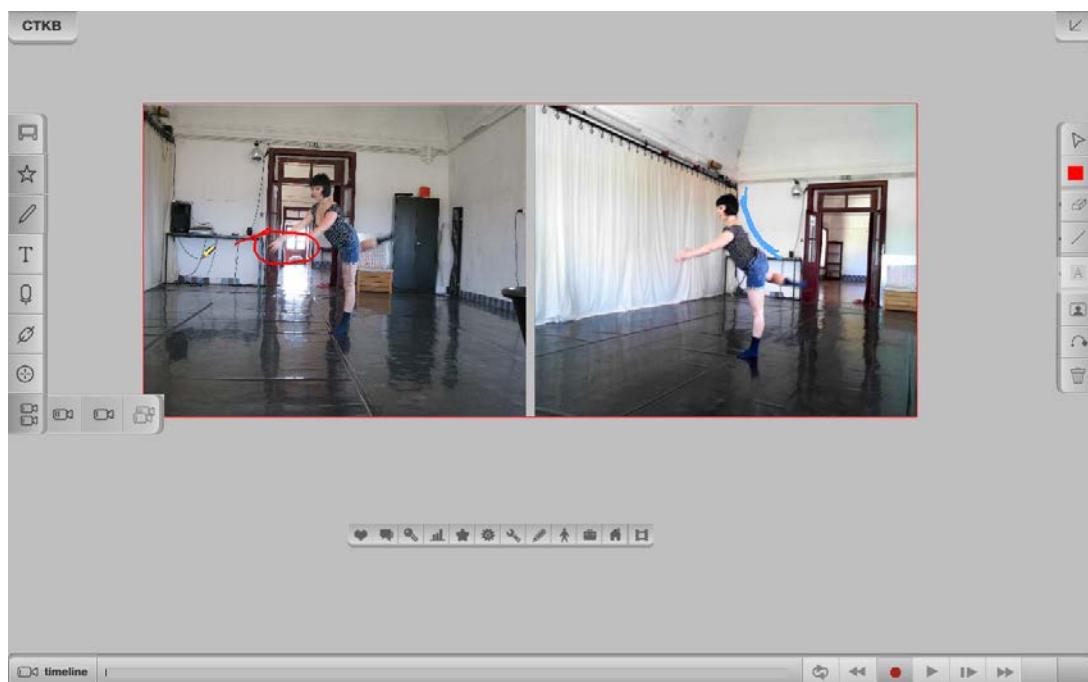
A solução permite capturar até duas *streams* de vídeo, provenientes de duas fontes diferentes. As fontes são dispositivos que captam e enviam vídeo para a ferramenta, como ecrãs, câmaras de vídeo, câmaras web ou projectores. Tal variedade de inputs, dá resultado a novas experiências e abordagens ao trabalho em causa. Por exemplo, esta funcionalidade torna exequível uma instalação multi-câmara, o que controla e admite uma análise mais profunda sobre o motivo que está a ser capturado. Outro exemplo é utilizar como fontes uma câmara e um vídeo proveniente de um ecrã, em que o objecto da cena, que está a ser capturado pela câmara, interage com o conteúdo do vídeo, resultando numa experiência multimédia interactiva.

A visualização das *streams* de vídeo pode ser individual ou simultânea, ou seja, é possível visualizar o vídeo de uma fonte ou os vídeos das duas fontes, dispostos um ao lado do outro. Neste último caso e sendo vid1, o vídeo localizado no lado esquerdo, e vid2, o vídeo situado no lado direito,

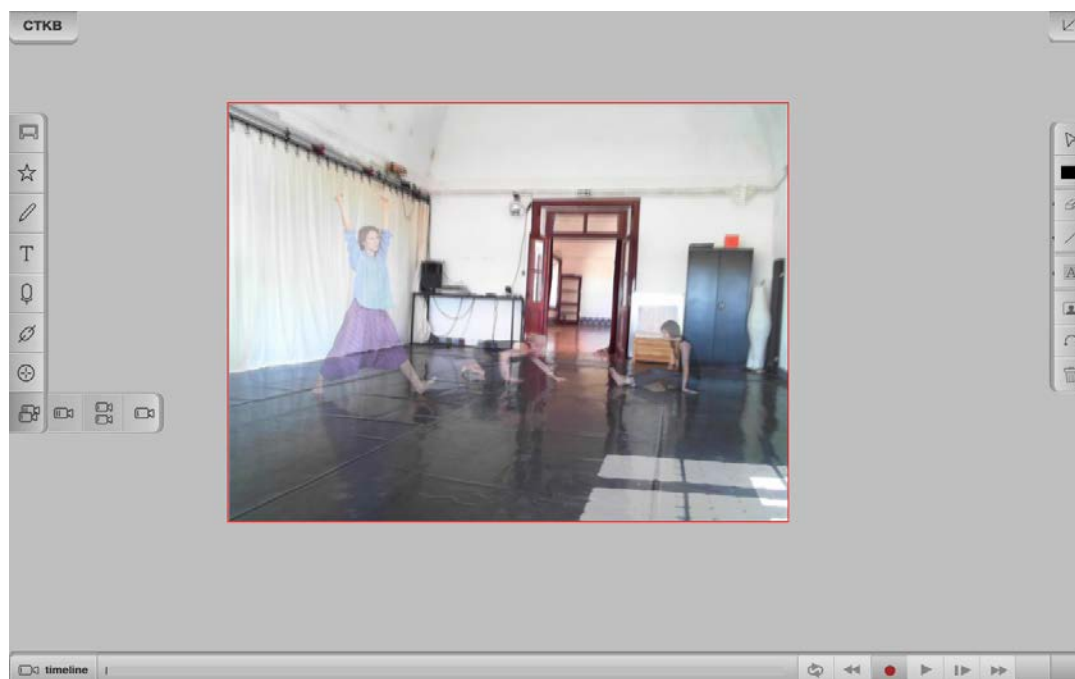
as anotações inseridas são associadas ao vid1, se a sua localização espacial estiver ao lado esquerdo do vid2 e são associadas ao vid2 se estiverem situadas ao lado direito de vid1, como explica a figura 3.4. Também existe a possibilidade da visualização do vídeo de uma fonte estar sobreposta à visualização do vídeo da outra fonte através de camadas, em que a camada superior é semitransparente, como mostra a figura 3.5.

Para iniciar a captura de vídeo, é necessário activar a visualização das câmaras. Para isso, pressiona-se o botão *camera tool*, que apresenta o vídeo proveniente da fonte seleccionada. Dentro deste botão, existe um sub-menu que permite visualizar o vídeo da fonte 1, da fonte 2, da fonte 1 e 2 colocadas uma ao lado da outra ou das 2 fontes sobrepostas, os quais podem ser mudados durante a captura. Para começar a captura, pressiona-se o botão *record* presente nos botões de reprodução do vídeo na *timeline*. Quando se encontra maximizada, a *timeline* apresenta o intervalo de tempo das anotações criadas ao longo do tempo, em forma de rectângulo.

Por questões de eficiência, existe a possibilidade de definir se a ferramenta captura apenas uma fonte ou captura as duas fontes. Se estiver configurada para capturar duas fontes de vídeo, é possível visualizar o vídeo de cada fonte individualmente ou simultaneamente durante a captura, mas ambas são sempre capturadas. Por omissão, os vídeos são gravados a uma resolução de 640 por 480 pixéis no formato *Quicktime (.mov)*.



**Figura 3.4:** Visualização simultânea das duas *streams* de vídeo. A anotação de tinta digital no lado esquerdo foi realizada ao lado de vid2, por isso é associada a vid1. A anotação de tinta digital do lado direito é associada a vid2 por esta estar desenhada à direita de vid1. Este tipo de visualização permite, por exemplo, uma apreciação mais pormenorizada sobre a performance do bailarino. Neste caso, é possível visualizar um movimento através de duas perspectivas diferentes: no lado esquerdo, pode-se observar a posição das mãos e braços e, no lado direito, a postura das costas.



**Figura 3.5:** Visualização das duas *streams* de vídeo sobrepostas. Este tipo de visualização pode ser conveniente para testar eventuais cenários na coreografia. Neste exemplo, a ferramenta está a capturar um vídeo da performance dos bailarinos (camada semitransparente) e a capturar outro vídeo de um cenário estático (camada opaca).

### 3.3.3.2 Modos de Anotação de Vídeo

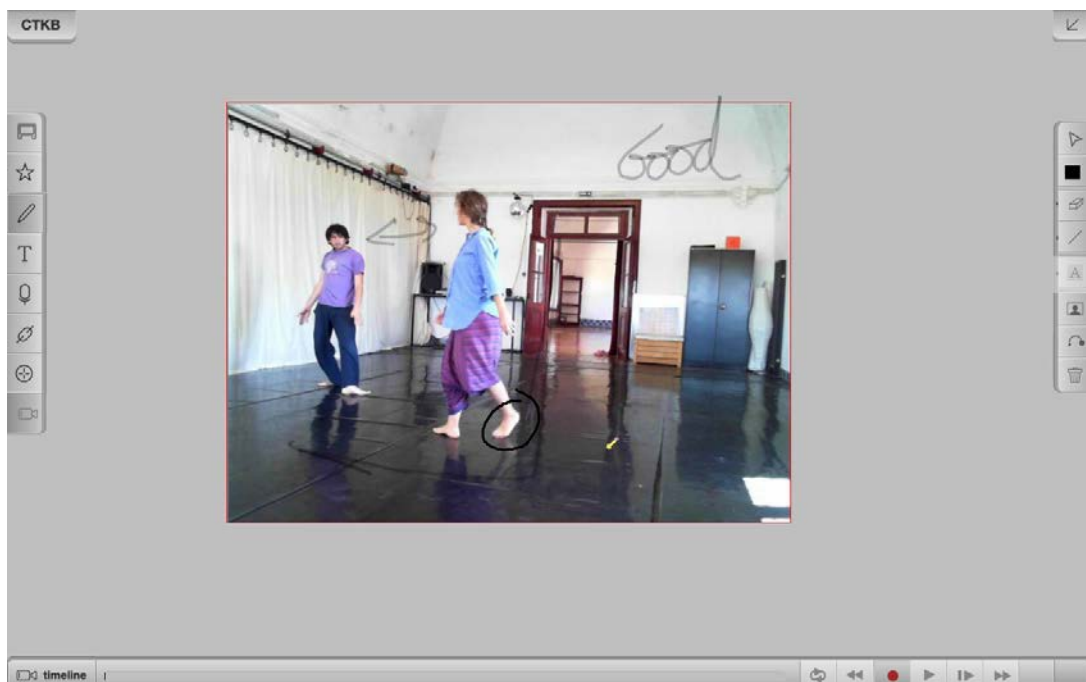
Para contemplar o requisito de tentar criar mecanismos para facilitar a percepção e anotação de momentos relevantes durante a captura, foram desenvolvidos três modos de anotação: o modo contínuo, modo atrasado e o modo suspenso. Estes modos foram criados para servir de suporte ao processo de anotação em tempo real, especialmente na associação de anotações aos momentos certos do vídeo, independentemente da duração dos mesmos.

O modo contínuo e o modo atrasado implicam duas metodologias de trabalho distintas durante a captura da cena. O modo contínuo é destinado a utilizadores que se concentrem mais na ferramenta e menos na cena ao vivo, contrariamente ao modo atrasado que se destina a utilizadores que normalmente foquem mais a sua atenção na cena ao vivo do que na ferramenta. Cada um destes modos tenta satisfazer as necessidades de cada utilizador durante o processo criativo, tendo em conta o seu método de anotação e o trabalho envolvido. O modo suspenso é um modo auxiliar que pode ser usado dentro do modo contínuo ou atrasado e serve para suspender a captura de modo a anotar um instante específico. Estes modos são acedidos através do botão *annotation mode* localizado na *tool bar*.

#### Modo Contínuo

No modo contínuo (*continuous mode*) ilustrado na figura 3.6, que é o modo por omissão, as anotações são criadas enquanto o vídeo é capturado, por isso são gravadas ao longo do vídeo em segmentos.

Este modo permite ao utilizador visualizar o vídeo a ser capturado continuamente, em tempo real, e aplicar as suas anotações. As anotações criadas neste modo são também contínuas, no sentido em que, quando inseridas, gradualmente vão desaparecendo à medida que a captura avança. Neste modo, a metodologia de trabalho usada para anotação é mais focada no ecrã do Tablet PC e menos na cena ao vivo, pois a visualização da captura está sincronizada com a cena ao vivo, sem atrasos. Assim, caso o utilizador visualize primeiro a cena ao vivo e depois a anote na ferramenta, existe um atraso entre o início do momento e o início da anotação, que corresponde ao tempo em que o utilizador desvia a sua atenção da cena ao vivo para a ferramenta, mais o tempo que decorre até à inserção da anotação correspondente, o que tornaria a anotação dessincronizada com o segmento.



**Figura 3.6:** Modo contínuo - anotações desvanecem-se com o tempo.

### **Modo Atrasado**

O modo atrasado (*delayed mode*) funciona da mesma forma que o modo contínuo, mas a visualização da captura do vídeo na ferramenta encontra-se atrasada relativamente ao que se passa na cena ao vivo. Com este atraso, a metodologia de trabalho é diferente do modo contínuo: o utilizador observa primeiro a cena ao vivo e, se encontrar algum momento susceptível de anotação, foca a sua atenção na ferramenta, que mostra a mesma cena a acontecer outra vez, anotando-a com mais tempo e precisão. O tempo de atraso pode ser especificado pelo utilizador em segundos.

### **Modo Suspenso**

O modo suspenso (*suspended mode*) permite suspender a visualização da captura, para o utilizador ter a oportunidade de anotar um instante representado por uma *frame*. A captura de vídeo em si não

suspende, continuando a decorrer normalmente em segundo plano. As anotações criadas são associadas a essa *frame* particular e, contrariamente aos outros modos, não desaparecem do ecrã até escolha de uma nova *frame* para anotação ou mudança de modo. Para visualizar as anotações após a captura, a reprodução de vídeo suspende na *frame* anotada, durante um período. Na figura 3.7, está representada a interface deste modo, que é composta por duas áreas: do lado esquerdo está a visualização da captura a decorrer e do lado direito a *frame* escolhida para futura anotação. Para escolher uma nova *frame*, o utilizador pressiona a área da visualização da captura de vídeo dentro deste modo, e nesse momento, a *frame* correspondente é copiada para o lado direito, permitindo a criação de anotações. A visualização da captura do lado esquerdo pode estar com velocidade normal (modo contínuo) ou atrasada (modo atrasado) relativamente ao evento ao vivo, por isso este modo pode ser usado dentro dos restantes. Se o modo suspenso for escolhido enquanto o modo contínuo estiver activo, a velocidade da visualização da captura é normal. Caso seja escolhido enquanto o modo atrasado estiver activo, então a velocidade é atrasada.

Este modo serve para o utilizador anotar momentos efémeros que sejam importantes, como um movimento ou acontecimento rápido, e analisá-los em pormenor. Pode ser usado também para anotações mais elaboradas, já que as anotações criadas nos outros modos são geralmente mais rápidas e simples e desvanecem-se com o tempo durante a captura. A visualização da captura na área esquerda faz com que o utilizador não perca os momentos do vídeo enquanto está a analisar e anotar o instante escolhido.



**Figura 3.7:** Modo suspenso - anotações permanecem visíveis. No lado esquerdo, está presente a visualização da captura a decorrer e, do lado direito, a *frame* escolhida para futura anotação.

**Tabela 3.1:** Comparação entre os diferentes modos de anotação.

	<b>Contínuo</b>	<b>Atrasado</b>	<b>Suspenso</b>
<b>Conceito</b>	Visualizar captura sincronizadamente com a cena ao vivo	Visualizar captura com atraso em relação à cena ao vivo	Suspender visualização da captura
<b>Objectivo</b>	Anotar continuamente	Anotar com atraso	Anotar instante
<b>Método de Trabalho</b>	+ Ferramenta - Cena ao vivo	+ Cena ao vivo - Ferramenta	Depende do modo (contínuo ou atrasado)
<b>Comportamento Gráfico das Anotações</b>	A desvanecer	A desvanecer	Visíveis
<b>Associação de Anotações</b>	Segmento	Segmento	<i>Frame</i>

### 3.3.3.3 Tipos de Anotação de Vídeo

Para usufruir dos vários inputs disponíveis para entrada de dados e acrescentar uma variedade de tipos de informação multimédia, foram criados quatro tipos de anotações de vídeo que podem ser aplicados: tinta digital, texto, marcas e áudio. Quando são usados em conjunto, tornam-se num material rico que acrescenta informação completa, em forma gráfica e sonora. De seguida, são apresentados cada um dos tipos de anotação, no que diz respeito ao seu conceito, propósito e forma de inserção. No final, a tabela 3.3 apresenta um resumo de todos os tipos de anotação de vídeo existentes na ferramenta. Pode-se dizer que, pelos seus métodos rápidos e simples de inserção, os tipos de anotação tinta, áudio e marcas são, em teoria, os tipos mais adoptados durante a captura.

### Anotações Gráficas

As anotações gráficas existentes são as anotações de tinta digital, texto, hiperligações e marcas. Ao contrário do áudio, estas têm uma representação gráfica, visível. Para inserir notas com base nestes tipos, usa-se a caneta digital ou os dedos. Cada anotação deste tipo tem uma localização associada, que pode ser dentro ou fora do conteúdo do vídeo.

#### Tinta Digital

A solução implementa anotações baseadas na caneta digital, isto é, um conjunto de traços de tinta com dimensões temporais e espaciais. As anotações de tinta servem para desenhar ou escrever texto livremente no vídeo, sendo o tipo de anotação que mais estimula a criatividade do utilizador. Cada



anotação de tinta é associada ao vídeo a partir do momento em que, com a ferramenta de tinta activada, o utilizador pressiona o ecrã para começar a desenhar o traço e deixa de o ser no momento em que a caneta deixa de fazer contacto com o ecrã, isto é, quando o traço acaba de ser desenhado.

Para activar este tipo de anotação e modificação das suas propriedades, basta carregar no botão *ink tool* na *tool bar*. É possível mudar a cor e espessura da linha acedendo para tal ao menu das *toolsets*.



**Figura 3.8:** Inserção de anotação de tinta digital.

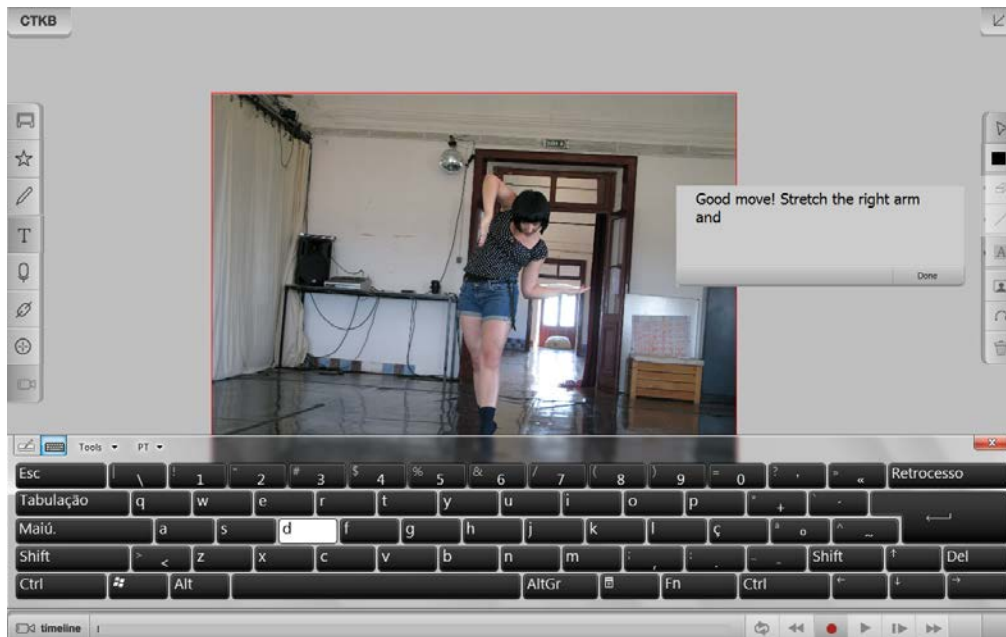
## Texto

Existe a possibilidade de inserir anotações de texto, escritas através de um teclado físico ou virtual, ambos disponíveis no Tablet PC. Este tipo de anotação é proveitoso, pelo facto de oferecer uma forma de escrever comentários textuais que sejam legíveis pela máquina, o que torna a procura deste tipo de anotações simplificada. O seu objectivo é, essencialmente, dar a oportunidade ao utilizador de escrever comentários extensos ou mais detalhados que complementem a informação de outras anotações, por isso a inserção deste tipo de anotação pode ser um processo demorado, que pode requerer mais tempo e concentração do utilizador. O facto de escrever as anotações de texto com o auxílio do teclado também contribui para que este tipo de anotação precise de uma disponibilidade acrescentada do utilizador para escrever, logo, as anotações de texto são maioritariamente inseridas em pós-captura.

Para criar uma anotação de texto, o utilizador activa a ferramenta de anotação de texto no menu *tools* e pressiona no local onde deseja escrever. De seguida, é apresentada uma caixa de texto no mesmo local, onde o utilizador pode inserir o texto pretendido. Se o portátil estiver no modo Tablet PC, é disponibilizado também o teclado virtual presente no sistema operativo (Microsoft Windows

Tablet PC Input Panel), por baixo da caixa de texto. O Tablet PC Input Panel fornece um teclado virtual de tamanho ajustável, juntamente com um reconhecedor de texto escrito à mão usando a caneta digital. Portanto, além do teclado, a caneta digital também pode ser um meio para se inserir texto. Quando o utilizador acabar de escrever o texto correspondente à anotação, carrega no botão *Done*, disponível na parte inferior da caixa de texto, para definir o fim do intervalo da anotação.

A cor, tamanho e fonte de letra do texto podem ser definidas, através dos botões respectivos no menu *toolsets*.

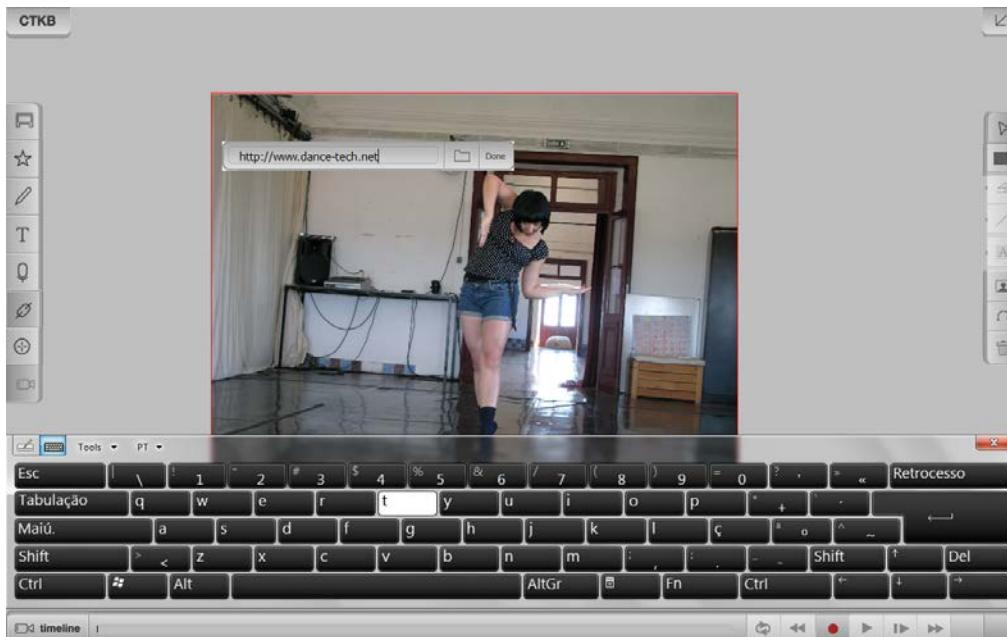


**Figura 3.9:** Inserção de anotação de texto no modo Tablet PC.

### Hiperligações

As hiperligações (*links*) são compostas por texto que têm uma ligação local ou remota a um objecto. O objecto pode ser um web site (ligação remota) ou o caminho de um ficheiro residente na máquina (ligação local). Os tipos de ficheiros locais podem ser de som, imagem, vídeo ou texto.

Uma anotação deste tipo é criada com a activação da ferramenta de hiperligação (*link tool*), que apresenta o teclado virtual e uma caixa composta por uma linha de texto. Aqui o utilizador pode escrever o URL de um web site ou procurar ficheiros locais, pressionando o botão à direita da caixa. O botão dá acesso ao navegador local de ficheiros que dá a oportunidade de escolher ficheiros multimédia que sejam suportados pela ferramenta, pois eles são apresentados ou reproduzidos dentro da mesma. O conteúdo do site também é apresentado numa janela incorporada na ferramenta, possibilitando a sua navegação. Similarmente às anotações de texto, quando se carrega no botão *Done*, é marcado o fim do intervalo da anotação.



**Figura 3.10:** Inserção de anotação na forma de hiperligação.

## Marcas

As marcas (*bookmarks*) são conceitos definidos pelo utilizador, como “cena”, “corte” ou “bravo!”, representadas por um ícone. É uma palavra-chave visual que define um conceito, emoção ou crítica. Cada marca é composta pelo seu nome e um ícone.

O botão *marks tool* dá acesso a todas as marcas disponíveis na ferramenta. Quando este botão é pressionado, as marcas surgem por baixo da visualização da captura na forma de uma lista horizontal, em que cada uma é representada por um botão com um ícone associado, que se relaciona com o seu significado. Para adicionar uma marca ao vídeo, pressiona-se no botão da marca correspondente. Nesse momento, uma cópia do botão aparece no topo do vídeo, o que significa que a marca está inserida. É possível arrastar este botão para outra zona do vídeo, que vai definir a posição onde a marca irá aparecer. Para terminar a associação da marca, pressiona-se novamente o botão da marca correspondente na lista. Esta lista encontra-se em baixo da visualização do vídeo, para que não exista sobreposição das mãos do utilizador com o vídeo no momento em que o mesmo marca o início e fim da associação de cada marca.

Além das 22 marcas de sistema pré-definidas, existe a possibilidade de criar novas marcas personalizadas, inserindo o seu nome e um ícone. Deste modo, o utilizador pode criar as suas próprias anotações e aplicá-las dentro do projecto actual. Tal funcionalidade torna-se profícua, na medida em que o utilizador tem a liberdade de construir e adaptar o seu vocabulário de anotações a cada projecto, consoante a sua natureza e o seu âmbito.



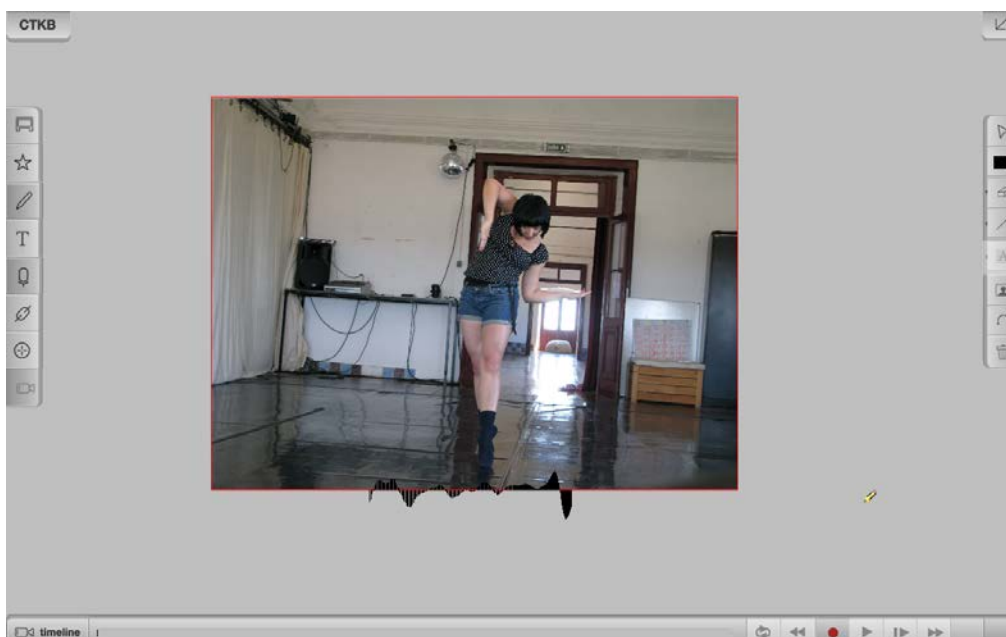
**Figura 3.11:** Inserção de anotação em forma de marcas com a barra de marcas aplicáveis em baixo.

### **Anotações de Áudio**

As anotações de áudio correspondem à gravação de voz do anotador e, portanto, são livres de gestos. São capturadas através de um microfone *Bluetooth*, para oferecer mobilidade ao utilizador no seu processo de anotação. O microfone *Bluetooth*, juntamente com o controlo remoto da ferramenta, promove uma forma de anotar por voz totalmente remota.

Cada anotação de áudio é activada e desactivada no botão *audio tool* na *tool bar*. Quando o utilizador pressiona no botão a primeira vez, a captura é iniciada até o botão ser pressionado novamente. Os instantes da activação e desactivação da captura correspondem ao início e fim do segmento anotado respectivamente. Durante a captura da anotação de áudio, é apresentada uma onda de som por baixo da visualização da captura, para mostrar *feedback* do input ao utilizador. Quando a anotação termina, é guardada como um ficheiro de som no formato *.wav*. Além das anotações de áudio, a ferramenta também captura o som ambiente do vídeo e armazena-o noutra ficheiro de som com o mesmo formato.

Este tipo de anotação é transparente e independente, no sentido em que cada anotação de áudio pode estar a ser capturada enquanto o outro tipo de anotações está a ser criada. Tal torna possível que mais do que um tipo de anotação seja associado a um segmento durante a captura, complementando ainda mais a informação. Por exemplo, uma combinação de anotação de áudio com anotação de tinta faz com que o utilizador possa falar do que está a ver e, ao mesmo tempo, apontar, através de traços de tinta, para a localização do vídeo que se está a referir, tal como acontece no sistema VideoTraces [15].



**Figura 3.12:** Inserção de anotação em forma de áudio.

**Tabela 3.2:** Comparação entre os diferentes tipos de anotação.

	<b>Tinta Digital</b>	<b>Texto/Hiperligações</b>	<b>Áudio</b>	<b>Marcas</b>
<b>Forma</b>	Traços de tinta	Texto/ <i>Link</i>	Voz	Palavra-chave visual, ícone
<b>Propósito</b>	Desenhos, pequenos comentários textuais	Longos comentários textuais/Web site ou ficheiro local	Comentários de áudio	Marcar eventos
<b>Input</b>	Stylus	Teclado físico ou virtual, Stylus	Microfone <i>Bluetooth</i>	Stylus, Toque
<b>Uso Geral</b>	Durante a captura, Pós-captura	Pós-captura	Durante a captura, Pós-captura	Durante a captura, Pós-captura
<b>Ativação (ponto de entrada/saída)</b>	Implícita (contacto do stylus), Local	Explícita (botão), Local	Explícita (botão), Remota	Explícita (botão), Local

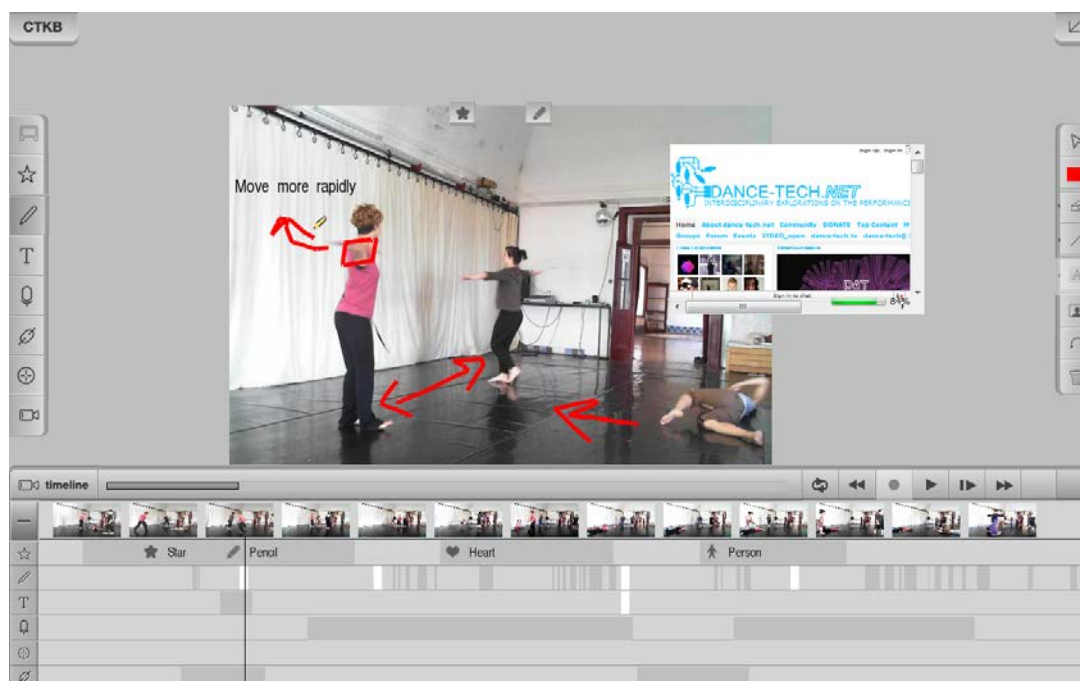
Os modos de anotação e os tipos de anotação tinta digital, áudio e marcas foram as soluções encontradas para solucionar o desafio da anotação em tempo real. Através dos modos de anotação, que tentam, na medida do possível, auxiliar o utilizador a não perder instantes da cena candidatos ao registo de informação e, através dos métodos de inserção rápidos e transparentes destes tipos de anotação, que não desviam a atenção do utilizador da cena, é possível ter todos, ou senão, a maioria dos momentos chave associados com informação personalizada no final da captura.

### 3.3.3.4 Navegação no Vídeo e Anotações

Os vídeos, capturados ou importados, e suas anotações podem ser navegados com o auxílio de uma *timeline* e dos botões de reprodução.

A *timeline*, apresentada na figura 3.13, é um gráfico que representa o tempo do vídeo por ordem, da esquerda para a direita. É composta por duas faixas principais: (1) uma faixa de treze *frames* do vídeo ao longo do tempo e (2) uma faixa de anotações, constituída por cinco pistas horizontais que correspondem aos tipos de anotação existentes (marcas, tinta, texto, áudio, *motion tracking* e hiperligações). Em cada uma das pistas, estão desenhados, em forma de rectângulo, os intervalos de tempo do tipo de anotação correspondente associados ao vídeo. É possível navegar no vídeo e anotações através da *timeline*, arrastando o cursor disponível que marca o tempo actual.

Os botões de reprodução localizados no topo direito da *timeline* permitem criar um *loop*, suspender, parar, avançar, recuar, reproduzir em velocidade normal ou em *slow motion* o vídeo. Estas operações também podem ser efectuadas manualmente arrastando o cursor da *timeline* na faixa de anotações.



**Figura 3.13:** *Timeline* e visualização de todos os tipos de anotação gráfica criadas. Na *timeline*, os rectângulos em branco correspondem a anotações associadas a uma *frame*, criada no modo suspenso e os rectângulos a cinzento correspondem a anotações associadas a um segmento, criadas no modo contínuo ou atrasado.

### 3.3.3.5 Edição e Visualização de Anotações

As anotações gráficas podem ser editadas após serem criadas. Para seleccionar uma anotação gráfica, usa-se a *selection tool*, no menu *toolsets*, para activar a ferramenta de selecção e, de seguida, pressiona-se na anotação no vídeo para a seleccionar. Além de ser possível alterar a posição da

anotação, movendo-a para o novo local, são permitidas três operações de edição, todas disponíveis no meu *toolsets*: agrupar, copiar e apagar. A operação agrupar consiste no agrupamento gráfico de duas ou mais anotações, sendo necessárias duas ou mais seleções. É usada sobretudo quando as anotações estão relacionadas por qualquer motivo e, por isso, devem-se comportar como um todo, como por exemplo uma anotação de tinta digital, em forma de seta, a apontar para uma anotação de texto. A operação copiar cria uma nova cópia da anotação e a operação apagar elimina-a definitivamente.

Além destas operações, é possível ajustar o instante inicial e final da anotação, recorrendo à *timeline* da seguinte forma: selecciona-se a anotação pressionando no rectângulo correspondente à anotação a editar, e executa-se um gesto para a esquerda ou para a direita conforme se deseje alterar o seu início ou fim respectivamente.

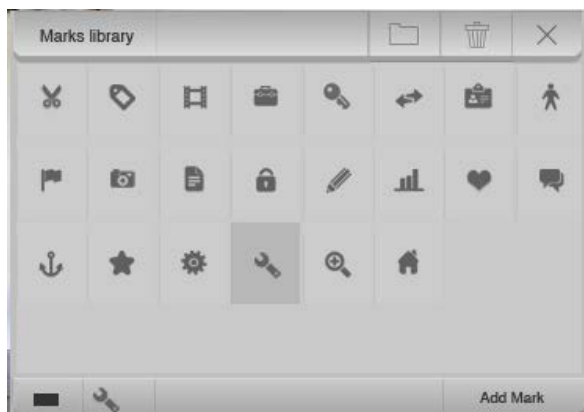
### 3.3.3.6 Gestão do Projecto

No separador *Project*, pode-se encontrar três submenus: os vídeos importados, as sessões capturadas e as faixas de anotação. Aqui estão todos os vídeos importados, que podem ser abertos com as respectivas anotações, as sessões capturadas, em que cada sessão corresponde a uma porção do vídeo que contém uma anotação associada e as faixas de anotação, que podem ser escondidas ou mostradas na *timeline*.

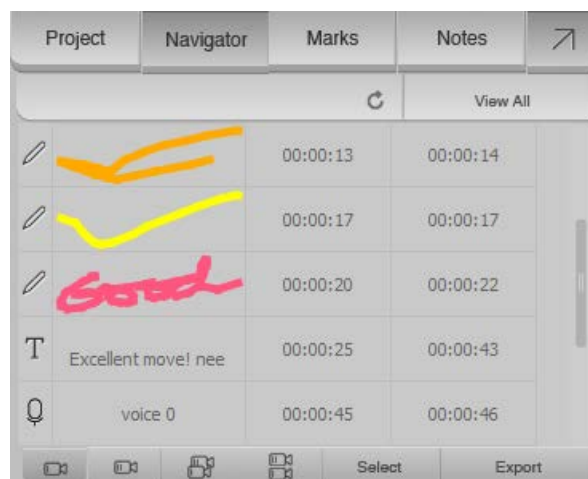
No separador *Navigator*, estão presentes, em forma de lista, o conteúdo de todas as anotações aliadas ao vídeo carregado e o instante inicial e final em segundos de cada uma, como apresenta a figura 3.15. O conteúdo das anotações de áudio é um nome pré-definido “audio note X”, onde X é um número sequencial. Se algum conteúdo da anotação for pressionado, o vídeo carregado avança ou recua para o instante inicial da respectiva anotação.

No separador *Marks*, são apresentadas, em forma de tabela, as marcas que estão a ser actualmente usadas no processo de anotação. Também se podem criar novas marcas, pressionando o botão *New Mark*. O *New Mark Menu*, ilustrado na figura 3.14, contém a biblioteca de todas as marcas disponíveis no projecto actual. Para inserir uma nova marca à biblioteca, é preciso inserir o nome da nova marca na caixa de texto e escolher o ícone correspondente. O ícone pode ser um de uma marca de sistema ou pode ser uma imagem externa. Se for uma imagem externa, o utilizador tem de pressionar o botão *Import Mark*, que dá acesso ao sistema de ficheiros da máquina para escolher a imagem. Quando é escolhida, a altura e largura da imagem são mudadas para 32 pixéis, de forma a se enquadrar dentro do botão. O utilizador pode ainda mudar a cor do ícone, pressionando o botão *Color Picker* situado à esquerda da caixa de texto. No final, ao pressionar *Add Mark*, a marca é adicionada ao conjunto das marcas que estão aplicáveis para anotação, presentes na barra e na tabela de marcas.





**Figura 3.14:** Menu *New Mark* que permite criar novas marcas personalizadas.



**Figura 3.15:** Menu *Navigator* que contém todas as anotações criadas com os seus instantes iniciais e finais.

No separado *Notes*, é possível escrever notas de texto livres, por exemplo, sobre o projecto em si. O texto pode ser ajustável quanto ao seu alinhamento, cor e fonte de letra. Estas notas são gravadas juntamente com o projecto e não estão relacionadas com nenhum vídeo, daí não serem consideradas anotações de texto no vídeo.



## 4. Realização

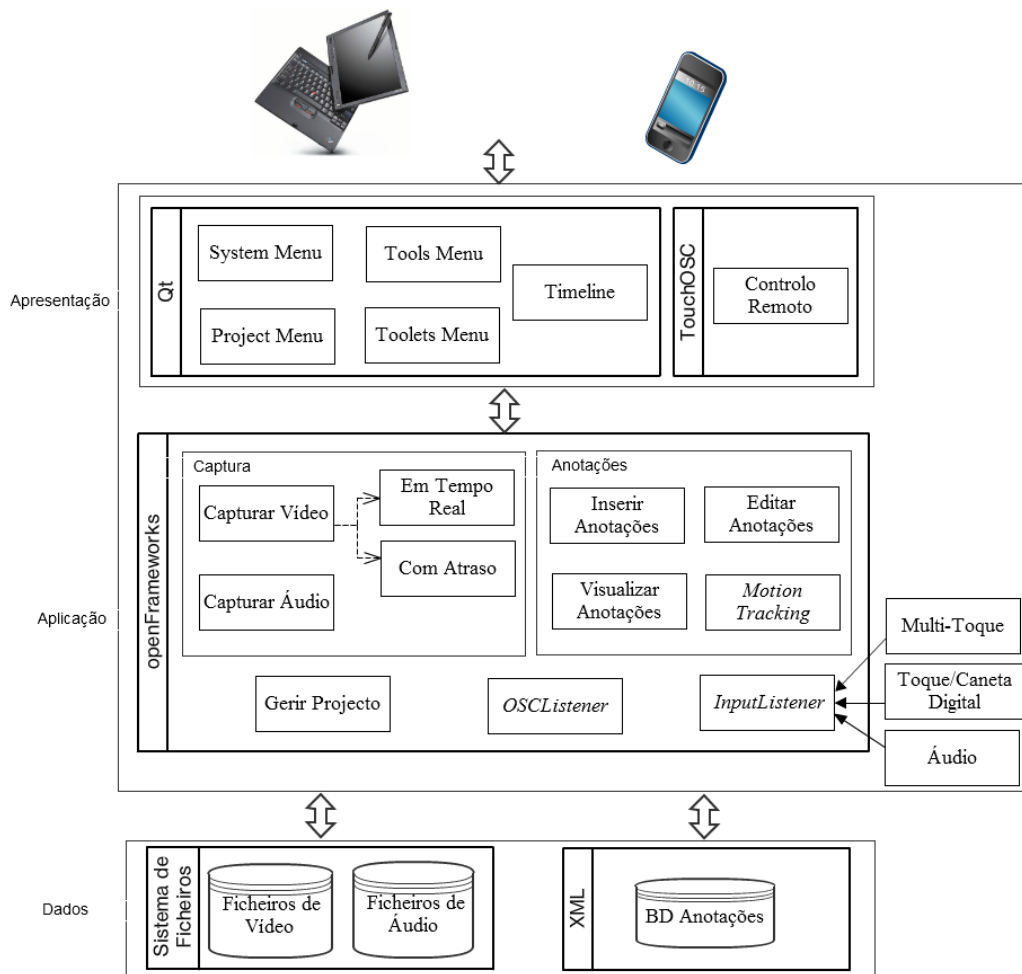
Este capítulo apresenta e descreve a arquitectura geral do sistema, bem como todos os seus componentes. Explica detalhadamente como foram realizadas as funcionalidades nas tecnologias propostas.

### 4.1 Arquitectura

O sistema implementado apresenta uma arquitectura em camadas, ilustrada na figura 4.1, e é constituída por três camadas principais: Dados, Aplicação e Apresentação. As camadas Aplicação e Apresentação estão contidas num mesmo módulo que representa a ligação e relação que existe entre elas. Estas camadas encontram-se separadas dentro do módulo, não só porque o seu conteúdo está desenvolvido em plataformas diferentes, mas também pelo facto da camada Apresentação ser construída no topo da camada Aplicação. Por exemplo, as janelas da interface da ferramenta no Tablet PC são construídas e sobrepostas à janela principal da camada Aplicação. Todas as subcamadas na arquitectura referenciam a plataforma onde foram desenvolvidas ou armazenadas, indicada no lado esquerdo.

A camada Apresentação é composta por duas subcamadas: a camada Qt, que contém todos os módulos correspondentes às áreas principais da interface do Tablet PC e a camada TouchOSC, que trata da criação da interface no dispositivo móvel. A camada Aplicação é constituída por duas áreas diferentes: a captura de vídeo e anotações de vídeo. Estas funcionalidades são tratadas de forma independente e, por isso, as anotações não são incorporadas no vídeo, tornando-se reutilizáveis. Além disso, esta camada também contém o módulo *OSCListener* que se destina a receber e computar mensagens OSC provenientes do dispositivo e o módulo *InputListener*, que visa distinguir os eventos detectados pelos dispositivos de entrada permitidos. Todos estes módulos foram desenvolvidos em openFrameworks. A camada Dados contém duas subcamadas que correspondem à subcamada dos

ficheiros de vídeo e áudio produzidos e armazenados e à subcamada da base de dados em XML de anotações de vídeo criadas e armazenadas.



**Figura 4.1:** Arquitectura geral do sistema.

Como referido na secção 4.3.1, o sistema implementado foi desenvolvido na linguagem C++, através da plataforma openFrameworks (OF) e Qt.

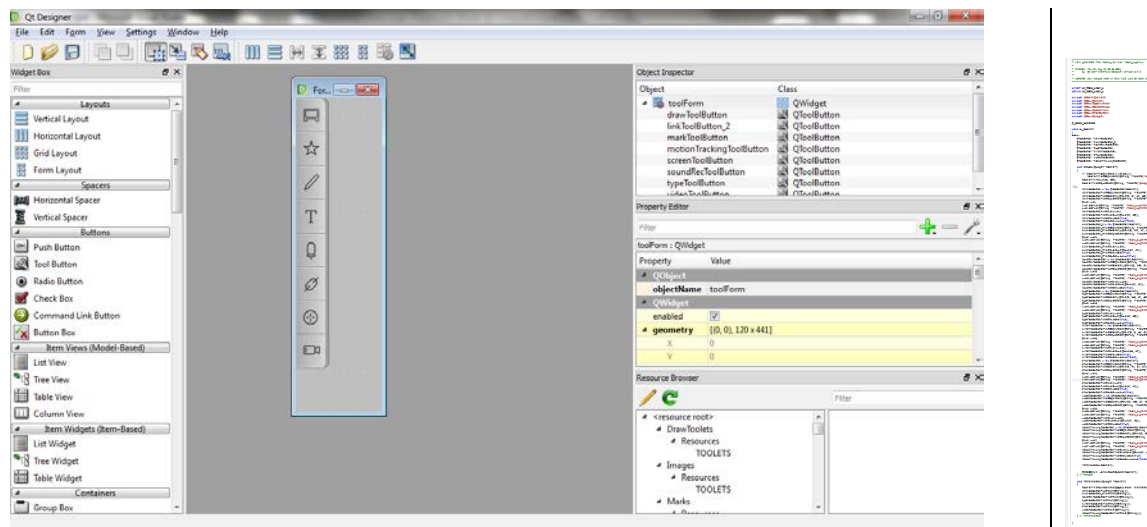


Figura 4.2: Qt Designer (área à esquerda) e respectivo código gerado (área à direita).

A plataforma Qt é composta por vários auxiliares para criação de interfaces, nos quais se inclui o Qt Designer. Esta aplicação permite construir interfaces graficamente e gera automaticamente o código resultante que as implementa.

Os elementos gráficos da interface são chamados de *widgets* e, por isso, cada *widget* deriva da classe *QWidget*, que por sua vez deriva da classe *QObject*. A classe *QObject* é a classe raiz que todas as classes do Qt herdam.

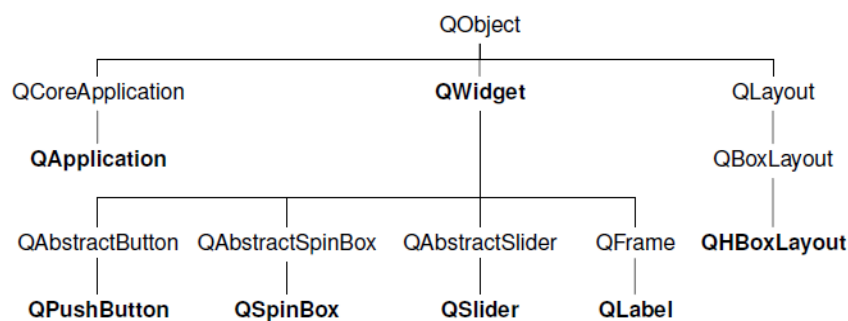


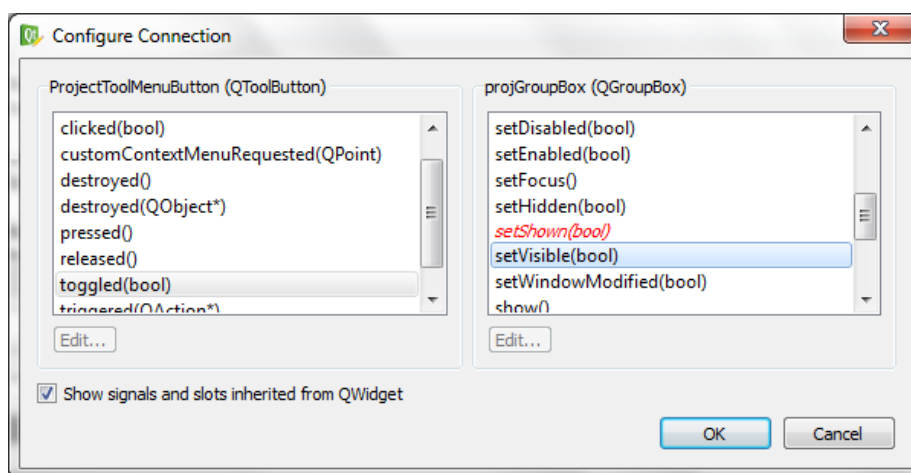
Figura 4.3: Árvore de herança de algumas classes Qt.

Tendo o desenho da interface implementada, é preciso combinar os sinais recebidos por cada elemento gráfico às acções que, em resposta, são desencadeadas. Para conectar uma ocorrência detectada num *widget*, como carregar num botão ou mover um *slider*, a uma funcionalidade, é usado o seguinte método:

**Listagem 4.1:** Declaração do método *connect*.

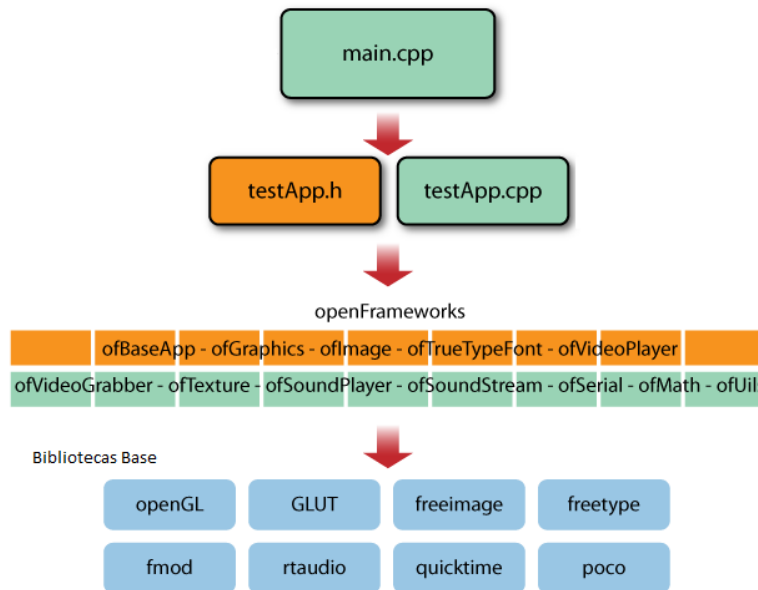
```
bool QObject::connect ( const QObject * sender, const char * signal, const QObject * receiver, const char * method, Qt::ConnectionType type = Qt::AutoConnection )
```

A função *connect* permite conectar um *signal*, que se trata de um evento emitido pelo *sender*, a um *slot*, que consiste numa função do objecto *receiver* com o código da funcionalidade esperada. Esta função é chamada sempre que um *signal* do *sender* for activado e retorna *true* ou *false* se a conexão tem ou não sucesso. Este é o método utilizado para relacionar as acções realizadas pelo utilizador na interface com as funcionalidades implementadas em OF.



**Figura 4.4:** Correspondência de *signals* de um *QToolButton* a *slots* de uma *QGroupBox*.

O OF encontra-se disposto em várias camadas de bibliotecas e código que usa essas bibliotecas, como ilustra a figura 4.5. O programa começa na classe *main.cpp*, que invoca a classe *testApp.cpp*, que contém o código do programa. Este código pode recorrer a uma série de classes disponibilizadas, construídas no topo das bibliotecas base, para lidar com som, imagem, vídeo, gráficos, entre outros.



**Figura 4.5:** Estrutura de uma aplicação em openFrameworks.

O núcleo de um programa em OF é a sua classe principal, *ofBaseApp*. Todos os métodos nesta classe são baseados em eventos. Isto significa que são desencadeados em resposta a eventos que acontecem dentro do programa, como tocar no ecrã, pressionar uma tecla ou desenhar os gráficos do programa. O cerne do OF trata de criar a janela e adicionar o código apropriado automaticamente de forma transparente.

Para se escrever uma aplicação em OF, apenas é necessário implementar os métodos herdados da classe *ofBaseApp*, entre os quais: *setup()*, *update()* e *draw()*. A classe *testApp* implementa estes métodos, sendo a classe principal da solução. O método *setup()* é invocado quando o programa inicia e serve para inicializar todas as variáveis e tudo o que seja importante de executar no início do programa. O método *update()* é chamado antes do método *draw()* e trata de processar os dados antes de serem desenhados. Finalmente o método *draw()* contém o código respectivo ao desenho dos gráficos do programa. Para além destes métodos, existem outros que invocam o seu código, quando detectam eventos executados provenientes dos periféricos de entrada de dados usado (*pressed*, *released*, *dragged*, *moved*).

## 4.2 Captura de Vídeo

Como especificado na secção 4.3.3.1, a ferramenta captura até duas *streams* de vídeo no formato *Quicktime* (*.mov*) a uma resolução de 640 por 480 pixéis.

Por razões de complexidade, o utilizador pode preparar a ferramenta para capturar uma ou duas fontes de vídeo. Se estiver configurada para capturar duas fontes, a ferramenta captura-as sempre,

mesmo que não estejam activadas para a sua visualização durante a captura. Se estiver adaptada para capturar uma fonte, a ferramenta só captura uma à escolha e ignora as outras se existirem. Consta-se que a captura de vídeo é a funcionalidade mais exigente em termos de processamento, por isso, a anotação em tempo real, particularmente a anotação de tinta digital, torna-se mais fluida, e consequentemente natural, se apenas uma fonte estiver a ser capturada. Contudo, é perfeitamente possível anotar agilmente quando duas fontes estão a ser capturadas.

#### 4.2.1 Abordagens Possíveis

Como foi referido anteriormente e agora explicado com mais detalhe, a tarefa de capturar um vídeo envolve três operações:

- 1) Captar as imagens provenientes da fonte de vídeo a um ritmo especificado.
- 2) Computar internamente, caso seja relevante, as imagens recebidas (transformações em termos de escala ou mudança de definição), e posteriormente adicioná-las a um objecto de vídeo idealmente ao mesmo ritmo anteriormente especificado.
- 3) Adicionar áudio sincronizado ao objecto de vídeo, o qual é gravado num sistema de ficheiros.

Para capturar o áudio, os passos são idênticos, mas em vez de imagens, são captados e capturados pacotes de áudio. Para implementar esta sequência de encargos, foram formuladas e analisadas várias hipóteses para perceber qual a melhor estratégia de capturar até dois vídeos sem comprometer a anotação em tempo real. Portanto, para se capturar e gravar um vídeo em disco, correspondente aos passos 2) e 3), existem três abordagens possíveis:

- a) **Persistente:** Adicionar cada imagem ao vídeo e gravar em disco.
- b) **Intervalada:** Gravar cada imagem em memória e, em instantes pré-definidos, transferi-las da memória para o vídeo e, seguidamente, para o disco.
- c) **Futura:** Gravar cada imagem em memória e, no final, transferi-las todas da memória para o vídeo e, seguidamente, para o disco.

Foram analisadas todas estas soluções, no que diz respeito à sua complexidade temporal e espacial. Como o objectivo é construir um vídeo que não tenha um limite de tempo pré-definido, a opção a) foi a escolhida para implementar esta função. A opção b) faz “congelar” a aplicação durante o armazenamento das imagens para o disco, o que faz com que o vídeo resultante produza paragens de reprodução nesses instantes. A opção c) tem uma alta probabilidade de causar a ocorrência de falta de memória, dependendo da duração da captura, por isso é claramente a pior opção. Verifica-se que

apesar de requerer algum processamento durante a captura, a opção **a)** é a opção mais viável e funcional tendo em conta a rapidez desejada na inserção de anotações em tempo real e o *framerate* adequado para o vídeo resultante.

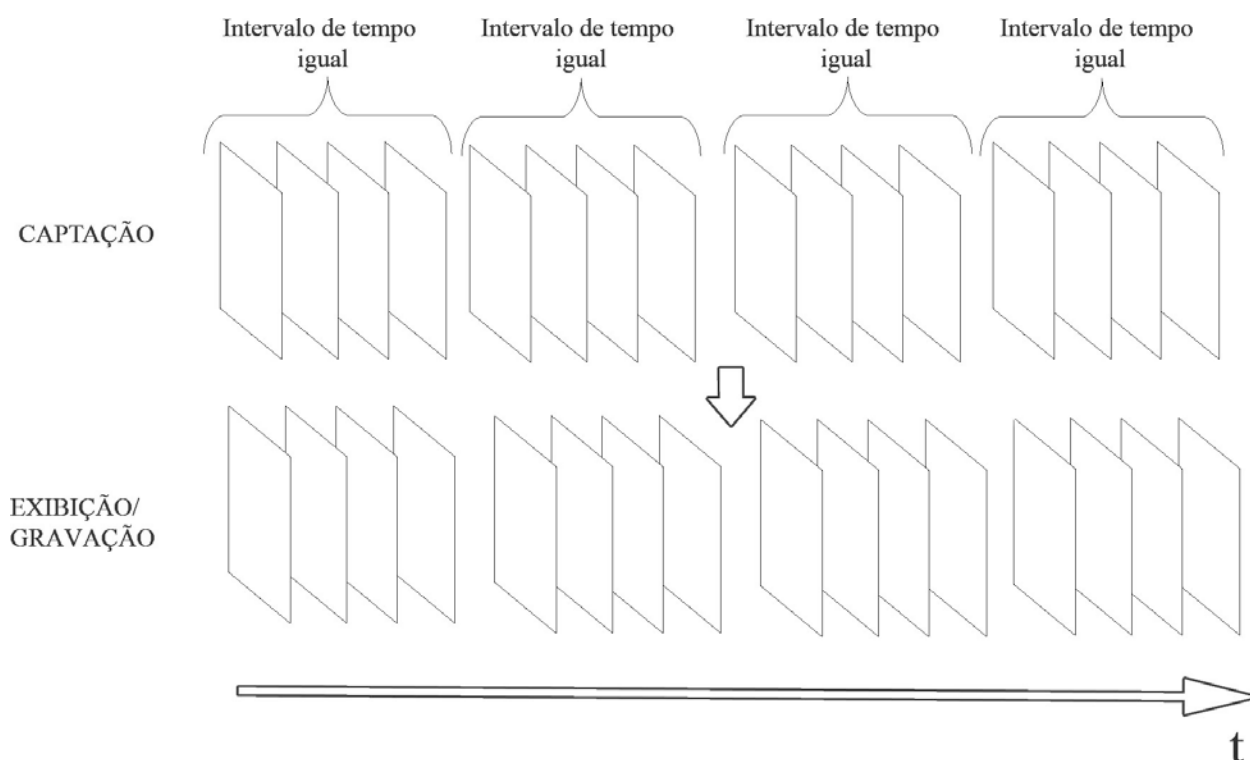
**Tabela 4.1:** Vantagens e desvantagens de cada tipo de captura quanto à carga de processamento exigida verificada.

	<b>Persistente</b>	<b>Intervalada</b>	<b>Futura</b>
<b>Carga de Processamento Durante a Captura</b>	Moderada	Elevada em instantes da captura, originando congelamentos do sistema.	Reduzida, até ocorrer falta de memória.
<b>Carga de Processamento no Final da Captura</b>	Reduzida	Reduzida	Elevada

#### 4.2.2 Modos de Anotação de Vídeo

Um factor importante a ter em conta ao implementar a captura de vídeo é que esta não devia recorrer a aplicações externas para a realizar. Sempre foi uma necessidade o facto da ferramenta não ter de depender de entidades exteriores. Assim, a captura de vídeo encontra-se programada com apoio da classe *ofVideoGrabber*, que capta as *frames* provenientes da fonte de vídeo, e da classe *ofxQtVideoSaver*, um *addon* que permite construir um vídeo com base nos pixéis das *frames* captadas. Ambas são disponibilizadas pelo openFrameworks e usam as bibliotecas do *Quicktime* para acederem e manipularem o conteúdo do vídeo.

No modo contínuo, as *frames* são adicionadas ao vídeo em tempo real. Portanto, quando uma *frame* é recebida vinda da fonte de vídeo, é adicionada instantaneamente ao vídeo resultante, como ilustra a figura 4.6. A *frame* é obtida da fonte através do método *ofVideoGrabber::grabFrame()* e é adicionada ao vídeo no método *ofxQtVideoSaver::addFrame()*, como comprova mais detalhadamente a listagem 4.2.



**Figura 4.6:** Exibição e gravação de *frames* ao longo do tempo no modo contínuo.

**Listagem 4.2:** Excerto do código correspondente à captura de vídeo em modo contínuo.

```

ofxQtVideoSaver vidSaver;
ofVideoGrabber vidGrabber;

void testApp::update(){
    vidGrabber.grabFrame(); // capta a frame
    vidSaver.addFrame(vidGrabber.getPixels(),1.0f/30.0f); // adiciona a frame ao vídeo a 30fps
}

void testApp::draw(){
    vidGrabber.draw(xGrabber, yGrabber, w, h); // apresenta a frame no ecrã
}

```

No modo atrasado, as *frames* são apresentadas e enviadas para o vídeo com um atraso especificado, como esclarece a figura 4.7. Para isso, é usado um *buffer* implementado numa fila, que guarda as *frames* recebidas à sua cauda. Quando o seu tamanho excede o seu limite, então é retirada a *frame* à sua cabeça, ou seja, a *frame* que está guardada há mais tempo, e é apresentada no ecrã. Desta forma, o vídeo é visualizado com um atraso que corresponde ao número de *frames* igual ao tamanho do *buffer*. Este processo é apresentado em forma de código na listagem 4.3.



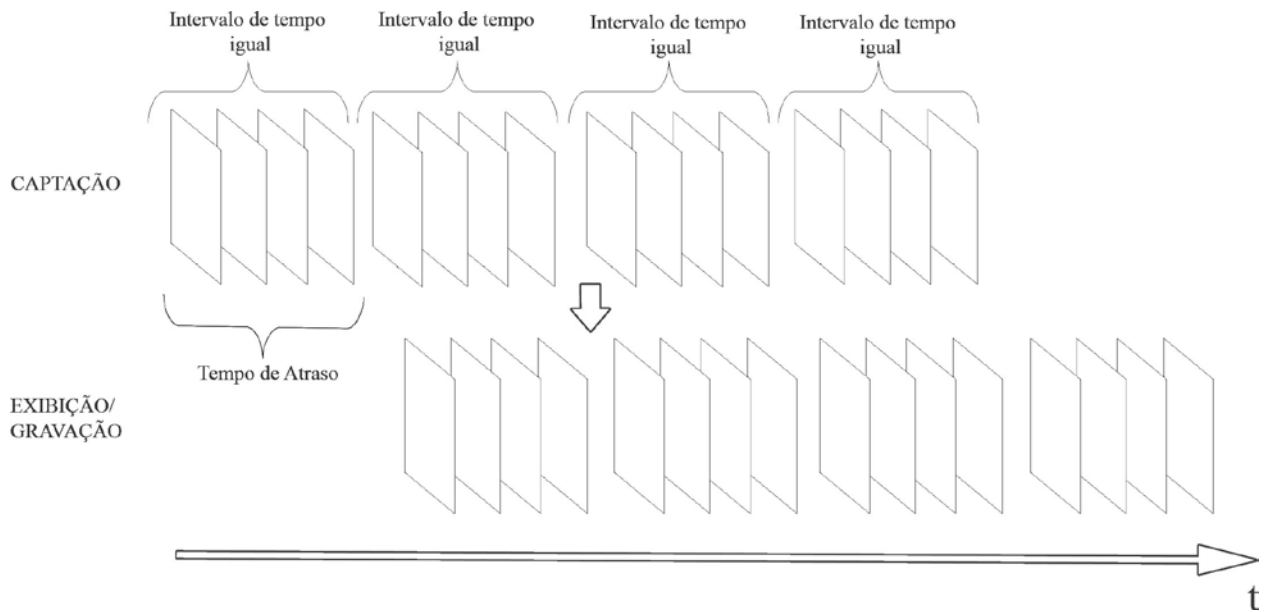


Figura 4.7: Exibição e gravação de *frames* ao longo do tempo no modo atrasado.

**Listagem 4.3:** Excerto do código correspondente à captura de vídeo em modo atrasado.

```

ofQtVideoSaver vidSaver;
ofVideoGrabber vidGrabber;
queue<ofImage> frames[20];

void testApp::update(){
    vidGrabber.grabFrame(); // capta a frame
    imageGrabber.setFromPixels(vidGrabber.getPixels(), CAMWIDTH, CAMHEIGHT,
    OF_IMAGE_COLOR ); // cria uma nova imagem contendo os pixéis da frame recebida
    vidSaver.addFrame(imageGrabber.getPixels(),1.0f/30.0f); // adiciona a nova imagem ao vídeo a 30fps
    frames->push(imageGrabber); // guarda a imagem no buffer
}

void testApp::draw(){
    if(frames->size() == 20 ){ // quando o buffer excede o seu limite (20 frames)
        frames->front().draw(xGrabber, yGrabber, w ,h); // apresenta a frame no ecrã
        frames->pop(); // retira a frame à cabeça
    }
}

```

O modo suspenso permite anotar um instante da captura, representada por uma *frame*. A visualização da captura é apresentada na área à esquerda e a *frame* escolhida para anotar na área à direita. Quando o utilizador pressiona a área à esquerda, uma nova imagem é criada contendo os mesmos pixéis da *frame* correspondente, a qual é apresentada na área à direita. Neste modo, o início e fim das anotações inseridas são iguais e correspondem ao instante da escolha da *frame* actual. Deste modo, são associadas ao instante especificado.

### 4.3 Captura de Áudio

As anotações de áudio e o som ambiente do vídeo são capturados, usando a biblioteca *libsndfile* v.1.0.25 e o *framework RtAudio 5*. *Libsndfile* é uma biblioteca escrita em C, que lê e escreve ficheiros com amostras de som, através de uma interface padrão. *RtAudio 5* está disponível no *openFrameworks* e serve de base à criação da classe *ofSoundStream*, que trata de captar e receber som originário da fonte de áudio. Existe um método baseado em eventos na classe *testApp*, chamado *audioReceived*, que é invocado automaticamente quando o som é percepcionado, como mostra a listagem 4.4.

**Listagem 4.4:** Excerto do código correspondente à captura de áudio.

```
void testApp::audioReceived(float * input, int bufferSize, int nChannels){
    if(videoRec && second == 0){
        // amostras estão "intercaladas"
        for (int i = 0; i < bufferSize; i++){
            leftAmbient[i] = input[i*2];
            rightAmbient[i] = input[i*2+1];
        }
        audioSaver.saveAmbientAudio(input, bufferSize);
    }
}
```

Este excerto de código captura o áudio ambiente do vídeo. Para capturar áudio das anotações, o processo é semelhante. O método recebe um apontador *input*, que referencia as amostras de som recebidas (valores entre -1 e 1), o tamanho das amostras *bufferSize* e o número de canais *nChannels*. Neste caso, o número de canais é igual a dois, ou seja em estéreo, por isso, as amostras de som recebidas encontram-se intercaladas. São guardadas em dois vectores diferentes (*leftAmbient* e *rightAmbient*) para as apresentar em forma de onda de som, de modo a dar *feedback* ao utilizador que o som está a ser obtido. A captura de áudio corresponde à última instrução do código, onde o apontador *input* é enviado para o objecto de áudio *audioSaver*, que com a ajuda do *libsndfile*, captura o som a uma taxa de amostragem definida e guarda-o num ficheiro *.wav*.

### 4.4 Anotações de Vídeo

Cada tipo de anotação de vídeo corresponde a uma classe diferente, em que cada classe foi concebida para guardar a informação específica a cada tipo. Cada classe correspondente a um tipo de anotação de vídeo herda da classe *Annotation*, que armazena os dados comuns de todos os tipos de anotação, incluindo o seu instante inicial e final e a fonte associada (1 ou 2), como mostra a figura 4.8. Portanto,

cada tipo de anotação é composto não só pelas suas propriedades exclusivas, presentes na sua classe respectiva, como também dos atributos iguais a todos os tipos de anotação herdados da classe pai. Desta forma, a inserção de novos tipos de anotação torna-se mais facilitada no futuro, devido à possibilidade de reutilizar código da classe raiz.

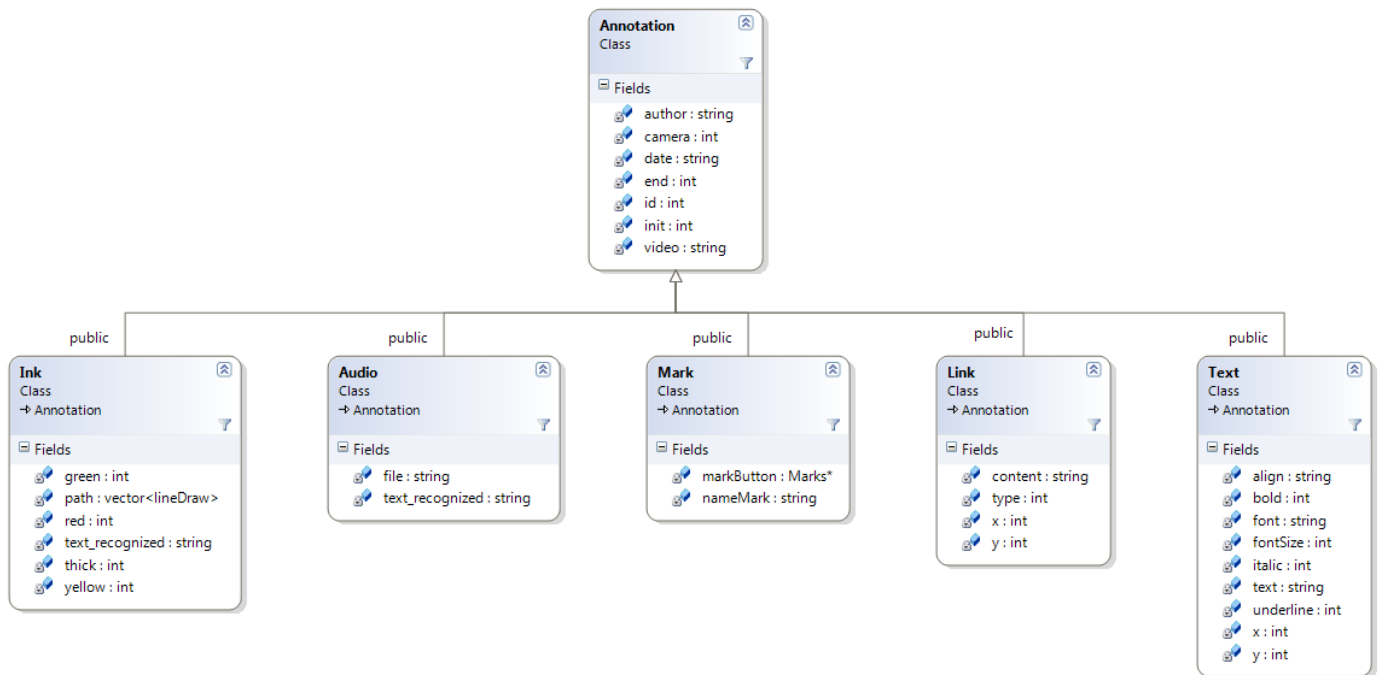
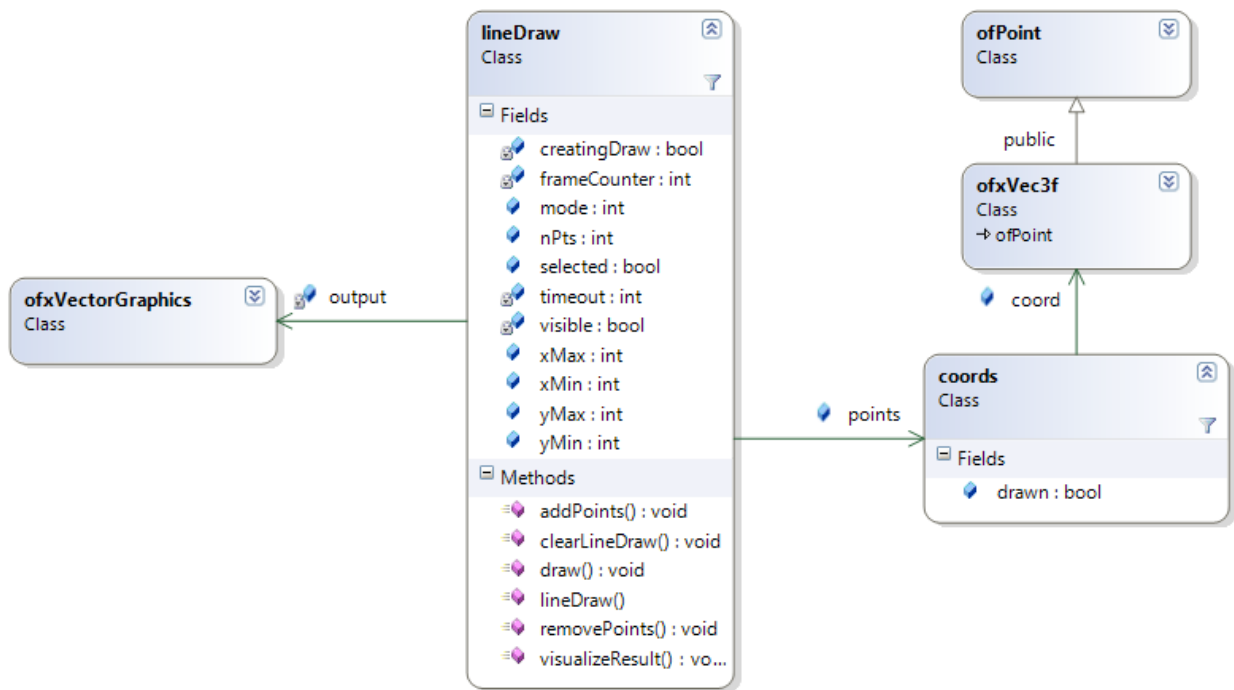


Figura 4.8: Resumo do diagrama de classes dos tipos de anotação.

#### 4.4.1 Tipos de Anotação de Vídeo

Cada anotação de tinta digital (*Ink*) é constituída pelas propriedades de formatação, como a espessura da linha e a cor, mas também pelas coordenadas do traço desenhado. Cada traço de tinta corresponde a uma anotação diferente e encontra-se guardado no vector *path*, que conserva em cada posição um objecto da classe *lineDraw*. A classe *lineDraw* é responsável pela gestão dos pontos da linha, como o seu desenho, adição e remoção. Por sua vez, os pontos da linha são preservados na variável *points*, um vector de objectos *coords*. A classe *coords* representa a coordenada do ponto (x,y), guardada na variável *coord*, e a sua visibilidade, guardada na variável *drawn*. A coordenada é do tipo *ofxVec3f*, que herda da classe *ofPoint*, ambas acessíveis pelo *openFrameworks*, que servem para armazenar e manipular pontos gráficos. Em suma, dentro de cada índice do vector *points*, estão presentes os pontos x e y herdados pela classe *ofPoint* e um booleano que identifica a visibilidade de cada ponto, isto é, se foi removido ou não, como ilustra a figura 4.9.

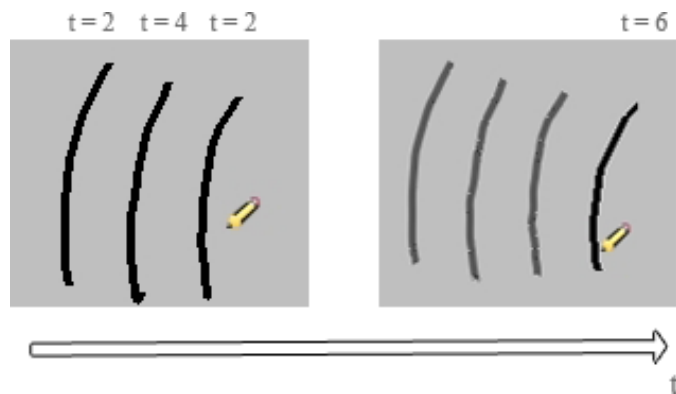


**Figura 4.9:** Resumo do diagrama de classes relativo ao tipo de anotação tinta digital.

O desenho das linhas que agregam os pontos adicionados é executado através da variável *output* do tipo *ofxVectorGraphics*, um *addon* fornecido pelo *openFrameworks* para desenhar gráficos vectoriais, implementado com recurso à biblioteca *OpenGL*.

Durante a captura, são desenhadas linhas conectando os pontos detectados no desenho efectuado pela caneta. Após a captura, em vez de linhas, são desenhadas curvas *Catmull-Rom* [27] compostas pelos pontos inseridos. As curvas *Catmull-Rom* são caracterizadas pela sua suavidade e continuidade  $C^1$ , que garante que não existe descontinuidade nas rectas tangentes à trajectória da curva. Assim, cada traço desenhado durante a captura é polido e regularizado, transformando-o num traço mais natural. Este melhoramento é realizado após a captura, para assegurar ainda mais fluidez ao desenho de traços em tempo real.

No modo contínuo ou atrasado durante a captura, quando o traço acaba de ser desenhado, desaparece gradualmente ao longo do tempo, isto é, aumenta a sua transparência até ao máximo. Tal acontece se o utilizador não desenhar outro traço a seguir ao anterior durante um tempo limite. Se desenhar um traço adicional a seguir ao anterior antes de se esgotar o tempo limite, o traço anterior permanece visível e o tempo limite é restabelecido. Se o tempo limite esgotar, então todos os traços anteriores visíveis gradualmente se tornam invisíveis e os seus instantes finais são igualados ao instante final do último traço. Esta particularidade permite ao utilizador esboçar e visualizar um desenho mais completo composto por vários traços de tinta, como exhibe a figura 4.10.



**Figura 4.10:** Exemplo do comportamento do desvanecimento dos traços de tinta, ilustrando um tempo limite  $t$  de 5 segundos. À esquerda, estão presentes três traços de tinta desenhados nos tempos limite  $t$  de 2, 4 e 2 indicados respectivamente. Quando um traço acaba de ser desenhado, o tempo limite  $t$  volta a 0. À direita, está presente um traço desenhado depois do tempo limite  $t$ , o que irá causar o desvanecimento dos traços desenhados anteriormente.

As anotações de texto, tal como as anotações de tinta, também têm propriedades gráficas ajustáveis, como o tamanho, cor e tipo da fonte. O seu conteúdo é texto, guardado na variável *text* do tipo *string*. A caixa de texto por onde se insere o texto, que vai constituir o conteúdo da anotação, é implementada em Qt instanciando a classe *QTextEdit*. Trata-se de um editor de texto que suporta formatação avançada usando tags HTML.

As hiperligações também são implementadas em Qt, usando a classe *QLineEdit*, um editor de texto de uma linha, mais simples que o edito da classe *QTextEdit*. No entanto, suporta operações como anular ou refazer uma acção ou cortar e colar texto. A variável *content* representa o conteúdo da hiperligação, isto é, o URL, e a variável *type* define se a ligação é local ou remota. Se a hiperligação é local, o ficheiro tem de ser do tipo som, imagem ou vídeo. O conteúdo é apresentado e reproduzido através das classes *ofSoundPlayer*, *ofImage* e *ofVideoPlayer* respectivamente, presentes no openFrameworks. Se a ligação for remota, o conteúdo do site inserido é apresentado dentro da ferramenta, com o auxílio do componente *QWebView* disponibilizado pelo Qt, que fornece um *widget* para ler e editar documentos web.

As marcas são botões criados em Qt, por meio da classe *QToolButton*. São compostas por um nome (*nameMark*) e o botão da marca (*markButton*). O botão é composto por um ícone, que consiste numa imagem relacionada com o nome da marca.

Quanto à visualização de anotações inseridas, todas as anotações são exibidas no mesmo intervalo de tempo no vídeo a que foram associadas. Para isso, no momento da associação e desassociação da anotação, é guardado o instante da primeira e última *frame* respectiva do intervalo na classe pertencente ao tipo de anotação inserido.

#### 4.4.2 Motion Tracking

A funcionalidade de *motion tracking* encontra-se em desenvolvimento no âmbito de outra dissertação, no entanto já se integrou parte desta funcionalidade na ferramenta apenas para teste e demonstração. Foram integrados dois *trackers*, o primeiro usando a técnica *FAST+BRIEF* e o segundo recorrendo ao dispositivo Microsoft Kinect<sup>10</sup>. O primeiro *tracker* faz com que uma anotação de tinta acompanhe o elemento de vídeo a que está associado, não obstante provoca a diminuição do *framerate* da aplicação, por isso, este *tracker* encontra-se em constante evolução. O segundo *tracker* detecta pessoas automaticamente com bastante eficácia e rapidez, desenhando um rectângulo à volta delas com uma cor diferente para cada pessoa, sem pôr em causa a aplicação. O problema deste *tracker* é que não necessita de nenhuma anotação para seguir um elemento no vídeo como acontece com o primeiro *tracker*, ou seja, não permite a escolha do elemento do vídeo que se deseja acompanhar, através da anotação. Assim, este processo também está a ser investigado, bem como a criação de novos *trackers* para futura integração na solução.

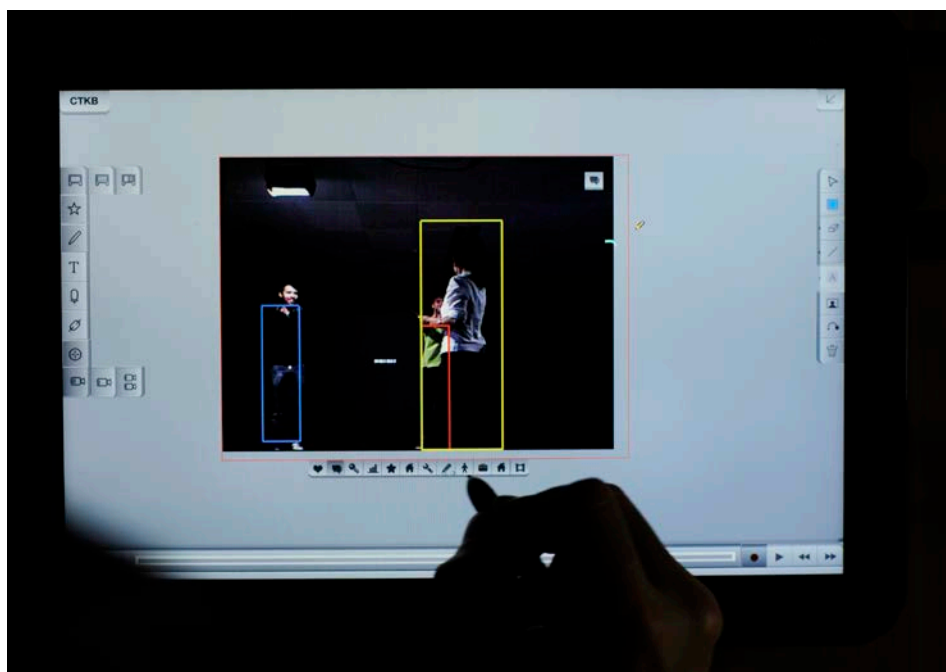


Figura 4.11: Motion tracking com recurso ao Microsoft Kinect.

#### 4.5 Armazenamento de Informação Gerada

A informação produzida pela ferramenta que constitui o output da solução é a criação de ficheiros de vídeo em formato *.mov*, ficheiros de áudio em formato *.wav* e eventuais anotações associadas em

<sup>10</sup> <http://www.xbox.com/pt-PT/kinect>

formato *.xml*. Cada conjunto de anotações associado a um vídeo é armazenado em forma de estrutura num ficheiro XML. A implementação teve como base o componente `ofxXMLSettings`, fornecido pela plataforma `openFrameworks`, que disponibiliza os mecanismos necessários para a utilização e manipulação de ficheiros XML pelas aplicações. Existe um módulo no sistema chamado `XMLUtils` que, recorrendo a este componente, trata de carregar e armazenar as anotações de vídeo criadas.

O documento XML é composto essencialmente por três nós principais, situados dentro da raiz do documento (*annotation\_document*):

- **metadata**: guarda a referência para o vídeo, bem como o nome do anotador e a data da última modificação.
- **annotation\_set**: armazena as anotações de vídeo, organizadas por tipo.
- **marks**: contém as marcas de sistema e marcas personalizadas disponíveis para aplicar.

Quanto aos tipos de anotação, todos os nós relativos aos mesmos contêm o seu ponto de entrada e saída, ou seja, a sua *frame* inicial e a *frame* final. Nas anotações de tinta digital (listagem 4.5), são armazenados os dados de formatação, através da *tag formatting*, que é composta pelos valores RGB da cor e a espessura do traço. A *tag <text\_recognized>* contém o texto reconhecido escrito com a caneta digital. Este reconhecimento não está implementado, mas está previsto como trabalho futuro. As coordenadas são guardadas dentro da *tag path*, composta pelas *tags* sucessivas `<x>` e `<y>`, que contêm cada coordenada *x* e *y* do traço armazenada. Esta estrutura pode ser observada em maior detalhe através do XML Schema apresentado na figura A.3, no anexo A.

**Listagem 4.5:** Excerto do ficheiro XML que representa o armazenamento de uma anotação de tinta digital.

```
<annotation_set>
  <annotation an:type="ink">
    <id>4</id>
    <text_recognized>Hello</text_recognized>
    <begin>245</begin>
    <end>678</end>
    <formatting>
      <color>
        <red>255</red>
        <blue>255</blue>
        <green>255</green>
      </color>
      <thickness>5.0</thickness>
    </formatting>
    <path>
      <x>23</x>
      <y>53</y>
      <x>35</x>
      <y>50</y>
    </path>
  </annotation>
</annotation_set>
```

Nas anotações de texto (listagem 4.6) são guardadas as propriedades de formatação do texto na *tag formatting*, isto é, cor, alinhamento, estilo, tamanho e cor da fonte. Além disso, também é guardado o texto, presente na *tag content*, a sua posição, identificada pelas *tags* *<x>* e *<y>* e o tipo de anotação de texto, guardado na *tag source*. Caso o tipo de anotação de texto seja 1, significa que se trata de uma anotação de vídeo, caso seja 2, é uma anotação textual relativa ao projecto. Esta estrutura pode ser observada em maior detalhe através do XML Schema apresentado na figura A.6, no anexo A.

**Listagem 4.6:** Excerto do ficheiro XML que representa o armazenamento de uma anotação de texto.

```
<annotation_set>
  <annotation an:type="text">
    <id>2</id>
    <source>1</source>
    <content>Teste anotação textual</content>
    <begin>345</begin>
    <end>1000</end>
    <x>300</x>
    <y>200</y>
    <formatting>
      <color>
        <red>255</red>
        <blue>255</blue>
        <green>255</green>
      </color>
      <align>left</align>
      <style>
        <bold>1</bold>
        <italic>0</italic>
        <underline>0</underline>
      </style>
      <size>12</size>
      <font>Verdana</font>
    </formatting>
  </annotation>
</annotation_set>
```

Nas hiperligações, é armazenado o tipo de hiperligação na *tag type*, que identifica se a ligação é remota (valor 1) ou local (valor 2). Se for remota, a *tag content* contém o URL do web site, se for local, contém o caminho para o ficheiro destinado. Esta estrutura pode ser observada em maior detalhe através do XML Schema apresentado na figura A.5, no anexo A.



**Listagem 4.7:** Excerto do ficheiro XML que representa o armazenamento de uma hiperligação.

```
<annotation_set>
  <annotation an:type="link">
    <id>5</id>
    <type>1</type>
    <content>http://www.dance-tech.net/</content>
    <begin>755</begin>
    <end>2303</end>
    <x>100</x>
    <y>800</y>
  </annotation>
</annotation_set>
```

As anotações de áudio (listagem 4.8) contêm a referência para o ficheiro de áudio capturado presente na *tag file*. Além disso, é guardado o texto reconhecido na *tag text\_recognized*. Esta funcionalidade também não está presente, mas estará no futuro, daí a razão pela qual esta *tag* existe. Esta estrutura pode ser observada em maior detalhe através do XML Schema apresentado na figura A.4, no anexo A.

**Listagem 4.8:** Excerto do ficheiro XML que representa o armazenamento de uma anotação de áudio.

```
<annotation_set>
  <annotation an:type="audio">
    <id>1</id>
    <text_recognized>Teste anotação áudio</text_recognized>
    <file>C:/Music/voz.wav</file>
    <begin>400</begin>
    <end>2000</end>
  </annotation>
</annotation_set>
```

Todas as marcas (de sistema e criadas pelo utilizador) disponíveis para serem aplicadas no vídeo são guardadas na *tag marks*. Cada nó corresponde a uma marca e é constituído pelo nome da marca e referência e cor do seu ícone. Tendo este grupo de marcas acessíveis para anotação, as marcas, de facto, inseridas no vídeo são armazenadas dentro da *tag annotation\_set*. Cada marca inserida contém uma referência, identificada pela *tag idref*, que aponta para a marca usada presente na *tag marks*. O conteúdo da *tag idref* é igual ao conteúdo respectivo da *tag id* da marca usada, como ilustra a listagem 4.9. Esta estrutura pode ser observada em maior detalhe através do XML Schema apresentado na figura A.1 e figura A.2, no anexo A.

**Listagem 4.9:** Excerto do ficheiro XML que representa o armazenamento de uma marca.

```
<marks>
  <mark an:id="1">
    <name>Luz</name>
    <icon>photo</icon>
    <color_icon>
      <red>12</red>
      <blue>231</blue>
      <green>123</green>
    </color_icon>
  </mark>
  <mark an:id="2">
    <name>Palco</name>
    <icon>scissor</icon>
    <color_icon>
      <red>255</red>
      <blue>231</blue>
      <green>123</green>
    </color_icon>
  </mark>
</marks>

<annotation_set>
  <annotation an:type="mark">
    <id>7</id>
    <idref>1</idref>
    <begin>50</begin>
    <end>2300</end>
    <x>20</x>
    <y>400</y>
  </annotation>
</annotation_set>
```

## 4.6 Controlo Remoto

*Open Sound Control* (OSC) é um protocolo de comunicação entre computadores, sintetizadores de som e outros dispositivos multimédia. Este protocolo é utilizado na solução com o objectivo de trocar mensagens entre o controlo remoto e a aplicação.

Actualmente, existem implementações variadas do protocolo adaptáveis para o contexto de cada aplicação. A classe *ofxOSC*, fornecida pelo *openFrameworks*, implementa este protocolo e foi utilizada para receber as mensagens provenientes do controlo remoto. Como especificado anteriormente, o controlo pode enviar para a ferramenta duas informações: a activação ou desactivação da captura de áudio e a activação ou desactivação da captura de vídeo.

As mensagens OSC são compostas principalmente por um *address* que identifica a mensagem, ou seja, uma *string* no estilo URL, e um argumento, que pode ser um inteiro, *float* de 32 bits, uma *OSC-string*, *OSC-blob* ou *OSC-timetag*. A aplicação *TouchOSC*, residente no controlo remoto e responsável por enviar mensagens para a ferramenta, apenas suporta o envio dos tipos inteiro e *float* através de UDP.

A implementação da comunicação recorre à classe *oscReceive*, fazendo com que a ferramenta, o servidor OSC, esteja constantemente em espera activa por mensagens provenientes do controlo

remoto, o cliente OSC. Este ciclo é executado dentro da função *osc.nextMessage()*. Como ilustra a listagem 4.10, quando uma mensagem é recebida na ferramenta, é verificada a sua *address* para identificar qual dos botões foi pressionado no controlo. Realizada esta distinção, é chamado o método respectivo para activação ou desactivação da funcionalidade. Esta activação ou desactivação é determinada pelo argumento recebido, que é um inteiro de valor 1.0 ou 0.0, que identifica se o botão no controlo foi pressionado para activação ou desactivação da funcionalidade respectivamente.

**Listagem 4.10:** Excerto do código correspondente à recepção e tratamento de mensagens OSC provenientes do controlo remoto.

```
oscReceive osc;

void testApp::setup(){
    osc.setup(); // inicializa as variáveis OSC
}

void testApp::update(){
ofxOscMessage* m = osc.nextMessage(); // em espera activa até receber nova mensagem
if(m!=NULL){
    string args = osc.getArg(m); // obtém o argumento
    if(osc.getAddress(m) == "/soundToggle") // obtém o address e compara
        enableDisableSound(args); // inicia/termina captura de áudio
    if(osc.getAddress(m) == "/recordToggle")// obtém o address e compara
        enableDisableRecord(args); // inicia/termina captura de vídeo
    }
}
```



## 5. Avaliação

Neste capítulo, apresenta-se o trabalho efectuado para a avaliação da solução implementada, através de testes de usabilidade e observação de utilizadores. Como este sistema é aplicado à área da dança, os utilizadores sujeitos à avaliação foram coreógrafos e alunos de dança contemporânea. Foi criado um inquérito para os avaliadores responderem, de forma a perceber melhor a utilidade, usabilidade e envolvimento emocional causado pela ferramenta.

Foram realizados quatro testes, dos quais os dois primeiros aconteceram na Faculdade de Ciências e Tecnologia, o terceiro no teatro Maria Matos (testes preliminares) e o último no Espaço do Tempo situado no Convento da Saudação em Montemor-o-Novo (teste final). Todos foram executados por coreógrafos de dança contemporânea, excepto o último teste, que foi feito pelos alunos de dança presentes no convento.

Os testes consistiram numa simulação de um ensaio de dança com um bailarino profissional, em que a ferramenta capturou a cena e o utilizador a anotou em tempo real. Para a captura da cena, recorreu-se a uma câmara web de alta definição (Microsoft LifeCam Studio) conectada ao Tablet PC e fixada num tripé posicionado em frente ao motivo a ser capturado. A outra fonte de vídeo utilizada foi a câmara web incorporada no Tablet PC. A ferramenta executou-se num Tablet PC HP TouchSmart tm2-2150ep com um iPod Touch 3G como controlo remoto e um microfone *Bluetooth* Nokia BH-105 para captura de áudio.

### 5.1 Testes Preliminares

Os três primeiros testes visavam não só examinar a usabilidade da ferramenta, no que diz respeito à sua qualidade e facilidade de utilização e aprendizagem, mas também produzir novos requisitos e ideias que podiam ser materializados em funcionalidades. Cada teste reuniu novas ideias sugeridas pelos utilizadores, o que permitiu moldar a ferramenta, deixando-a mais completa e multifacetada.

### 5.1.1 Descrição

O primeiro teste foi executado pelo coreógrafo Stephan Jürgens no dia 29 de Março, tratando-se do primeiro contacto exterior com a ferramenta. Nessa altura, existia apenas a versão *dark* da interface implementada e foram analisadas e averiguadas a captura e os tipos de anotação de vídeo tinta digital, texto e marcas. O segundo teste, tal como o primeiro, aconteceu na Faculdade de Ciências e Tecnologia, e os avaliadores foram os coreógrafos de dança David Santos e Carlos Oliveira. Ocorreu no dia 6 de Junho e serviu como uma preparação para o terceiro teste, que aconteceu três dias depois no dia 9 de Junho. Este terceiro teste foi o mais importante até então, já que aconteceu no teatro Maria Matos com a presença do conceituado coreógrafo Rui Horta. Aí, foi colectada e arquivada informação digital em forma de vídeo e fotos profissionais, tanto do ambiente do teste, como da experiência de utilização da ferramenta no Tablet PC. Com isto, passaram a existir materiais para demonstração da ferramenta. Nestes testes, a avaliação da ferramenta foi feita através do registo da experiência observada, no que diz respeito às dificuldades encontradas na interacção com cada funcionalidade posta à prova.



**Figura 5.1:** Primeiro teste preliminar com Stephan Jürgens realizado no dia 29 de Março.

### 5.1.2 Análise de Observações

Ao longo dos testes preliminares, a interface principal foi bem compreendida. Apesar disso, algumas dificuldades de interacção foram observadas, relacionadas com a capacidade de utilizar a caneta

digital e toque simultaneamente. Nomeadamente, quando o coreógrafo utilizava a caneta digital, o toque era bloqueado e vice-versa, contrariando uma verdadeira interacção bimanual. Contudo, esta limitação está relacionada com o hardware, isto é, o modelo do Tablet PC usado, que não permite a coincidência de aplicação destes dois dispositivos de entrada.

Na *timeline*, foram identificados problemas em interagir com o mecanismo da sua oclusão. Este mecanismo consistia em carregar no botão da *timeline* e de seguida arrastá-lo para cima ou para baixo, deslocando a *timeline*. Para se tornar mais fluido, este gesto foi simplificado, fazendo com que apenas o arrastamento do botão, sem carregar nele primeiro, cumprisse a função. Os botões de reprodução também foram melhorados, uma vez que se verificou que os botões de início e fim de gravação se tornavam mais familiares se fossem acoplados a um botão, em vez de dois separados, como anteriormente se encontravam. Durante a captura simultânea das duas fontes de vídeo, o desenho de anotações de tinta digital apresentou-se lento e, por isso, este aspecto foi aperfeiçoado, ao diminuir a sua complexidade e processamento desenhando linhas entre os pontos, em vez de curvas *Catmull-Rom*.

Quanto às marcas, constatou-se que a definição do fim da anotação distraía o coreógrafo da visualização da captura do vídeo, visto que, para executar tal operação, era necessário pressionar na marca inserida no topo da visualização do vídeo. Foi sugerido, e depois implementado, um segundo *click* na barra de marcas para marcar o instante final da anotação iniciada. O método de inserção das anotações de texto não apresentou dificuldades, mas foi relatado que este tipo de anotação seria mais conveniente para análise em pós-captura, uma vez que necessitam de mais tempo de input e atenção do utilizador.

Os modos de anotação foram bem recebidos pelos coreógrafos. Observou-se que foram bem percebidos e acessíveis. No entanto, foi sugerido que, nos modos contínuo e atrasado, as anotações gráficas permanecessem visíveis por mais tempo antes de se desvanecerem. Também se verificou que era necessário o desenho e implementação de outra versão da interface com um sistema de cores diferente, para a ferramenta se adaptar ao ambiente em que estava inserida.

Durante o desenvolvimento do sistema, o trabalho realizado foi frequentemente demonstrado a outras pessoas, como investigadores convidados para realizarem conferências na faculdade, colegas de mestrado e artistas ligados à dança. Nessas apresentações, foram reunidas novas ideias que se tornariam úteis para os coreógrafos, como a visualização do vídeo com atraso, de forma a que o anotador tenha disponibilidade para anotar a cena durante a captura do vídeo. Em resposta, foi criado o modo de anotação atrasado, que foi aplaudido pelos utilizadores alvo da ferramenta. Por isso, houve uma constante adaptação e implementação a novos requisitos durante o desenvolvimento, de modo a que correspondesse às expectativas dos utilizadores.

Depois dos resultados obtidos e respectiva análise adquirida nestes testes, foi feita uma reflexão sobre os aspectos mais críticos que mereciam uma nova abordagem. Adicionalmente, foram discutidas as novas sugestões recolhidas, quanto à sua utilidade e viabilidade no que diz respeito ao

tempo e recursos existentes na altura. Com estas metas definidas, foi marcado um teste final para se realizar um mês depois, que iria incluir todas as novas funcionalidades e melhoramentos para serem avaliados como um todo.

## **5.2 Teste Final**

O teste final realizou-se no dia 9 de Julho no Espaço do Tempo situado no Convento da Saudação em Montemor-o-Novo. Este teste foi presenciado pelo coreógrafo Rui Horta e toda a equipa pertencente a este projecto. O sistema testado já se encontrava num estado bastante avançado, uma vez que tinha sofrido várias iterações desde o início do seu desenvolvimento.

### **5.2.1 Descrição**

O objectivo do teste final não foi focado, contrariamente aos testes anteriores, em obter novos requisitos, mas sim em avaliar a usabilidade, utilidade e envolvimento emocional da ferramenta. Alunos de dança contemporânea foram alvo do teste, os quais experimentaram a ferramenta livremente e, no final, preencheram um pequeno inquérito sobre essa experiência. Não existiu um guião para o teste, no entanto cada aluno testou a funcionalidade principal da ferramenta: a captura e anotação de vídeo em tempo real, analisando todos os tipos e modos de anotação. Após essa utilização, poderiam usufruir abertamente das restantes funcionalidades, se houvesse tempo disponível para tal.

Os utilizadores foram divididos em 2 grupos, cada um com 5 pessoas: um grupo para testar de manhã e o outro de tarde. Cada teste ocorreu numa sala de ensaios de dança e demorou, em média, 30 minutos. Durante o teste, 4 utilizadores efectuaram uma performance em grupo ou individual, enquanto o restante testava a ferramenta. A câmara web de alta-definição, fixada num tripé, capturou a performance. Quando o avaliador acabava a sua experiência, um dos outros alunos começava uma nova avaliação.

No início do teste, foi disponibilizado um pequeno resumo sobre o contexto, âmbito e objectivo do projecto para uma melhor compreensão do seu enquadramento. De seguida, foi apresentada uma lista de tarefas que, no mínimo, teriam de ser realizadas durante o teste, isto é, anotar vídeo em tempo real com recurso aos vários modos e tipos de anotação. No final, o avaliador foi convidado a preencher um inquérito simples, disponível no anexo B. O coreógrafo Rui Horta também teve a oportunidade de experimentar a solução, o qual, quando todos os testes terminaram, se reuniu com a equipa de projecto para analisar e tirar conclusões sobre o estado da ferramenta.





**Figura 5.2:** Coreógrafo Rui Horta, em Montemor-o-Novo, interagindo com a ferramenta com a projecção da experiência atrás.

## 5.2.2 Inquérito aos Utilizadores

Com o objectivo de colher dados sobre usabilidade, utilidade e receptividade do sistema, foi realizado um inquérito anónimo de resposta rápida, para ser preenchido pelos utilizadores após cada teste. O questionário está dividido em duas partes, em que a primeira está relacionada com a avaliação geral da solução, através de classificações de utilização e acessibilidade das funcionalidades para uma análise quantitativa e através de sugestões ou comentários abertos para uma análise qualitativa. Estas classificações são escolhidas através de uma escala *Likert*. A segunda parte do questionário visa traçar o perfil do utilizador, nomeadamente obter informação sobre o seu método de trabalho e familiarização com as tecnologias presentes na solução, de modo a entender até que ponto a vivência passada influencia a interacção com a ferramenta.

O inquérito aborda o sistema em geral, mas foca principalmente os tipos e modos de anotação. As respostas às perguntas no inquérito permitem comparar cada tipo e modo, quanto à sua frequência de utilização e facilidade de aprendizagem. O inquérito está escrito em língua inglesa, pelo facto dos utilizadores alvo, os alunos de dança, não entenderem a língua portuguesa.

## 5.2.3 Análise de Resultados

Os resultados obtidos através do inquérito aos utilizadores são mostrados em maior detalhe no anexo B. Neste capítulo, é apresentado um resumo dessa análise, bem como as possíveis explicações e conclusões retiradas.

### 5.2.3.1 Utilizador

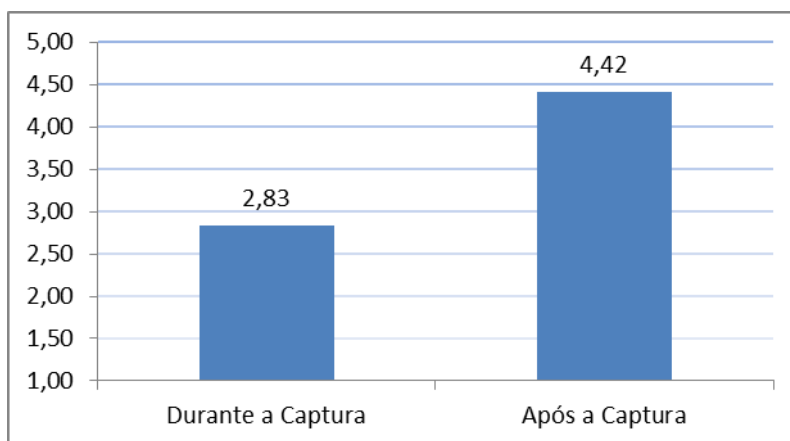
Foram inquiridos 12 utilizadores, dos quais 17% são do sexo masculino e 83% do sexo feminino, com idades compreendidas entre 19 e 27 anos, sendo que a faixa etária predominante se situou entre os 26 e 27 anos. O grau de instrução da maioria dos utilizadores é a licenciatura, sendo a habilitação académica com maior percentagem (58%). Dos restantes, 25% possuem o ensino secundário, 8% o ensino básico e outros 8% o grau de mestrado.

Quanto à familiarização das tecnologias baseadas em toque e baseadas em caneta digital, 17% e 8% respectivamente responderam que se encontravam habituados. Daqui se conclui que a interacção com a ferramenta constituiu uma experiência ainda mais nova e desconhecida para a grande maioria dos utilizadores.

Em relação ao processo de trabalho, todos responderam que habitualmente gravam o seu trabalho. O papel e o vídeo são os materiais mais populares para esse fim, sendo que 100% dos utilizadores captura vídeo e 92% anota em papel. Posto isto, a existência desta solução faz todo o sentido, já que passa por combinar estes dois tipos de materiais. Por um lado, captura o trabalho através do vídeo e, por outro, anota informação desse trabalho através da imitação da forma de anotar em papel. Além deste facto, 92% dos inquiridos responderam que costumam anotar durante o seu método de trabalho, usando na sua maioria papel (91%) e computador portátil (55%), o que reforça crescentemente a necessidade desta solução. A maioria dos utilizadores (75%) partilham o seu trabalho via e-mail e/ou hardware (CDs, DVDs, *pen drives*).

### 5.2.3.2 Cenários de Utilização

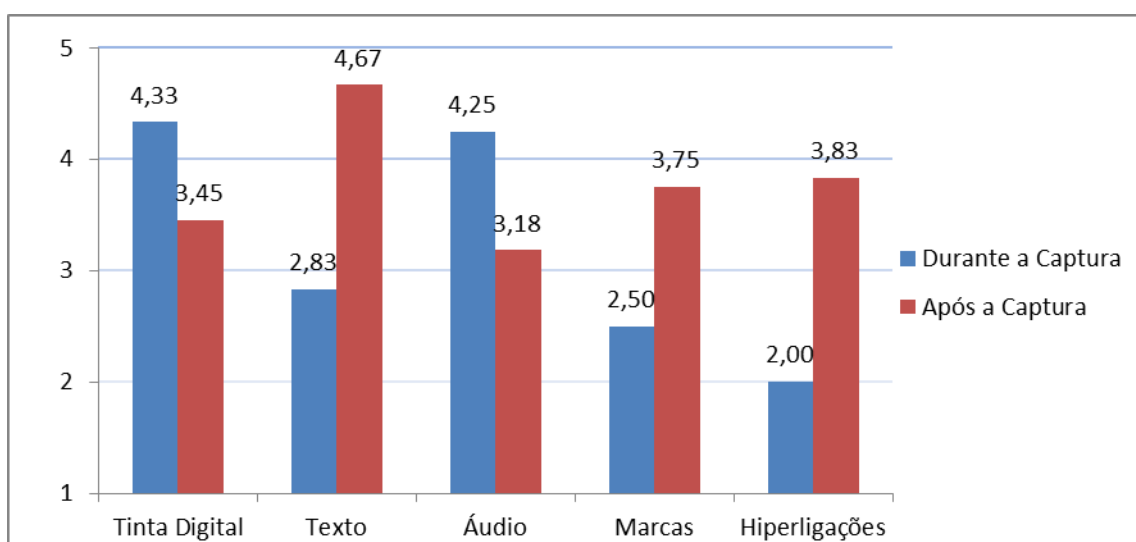
A 1ª pergunta do questionário visa conhecer qual o cenário mais apropriado para a aplicação da ferramenta, através da classificação pessoal da eventual frequência de utilização de cada um. A partir do gráfico 5.2, verificou-se que, em média, numa escala de 0 a 5, em que 0 significa “raramente” e 5 representa “frequentemente”, os utilizadores usariam a ferramenta, numa frequência de utilização de 2,83 durante a captura e 4,42 após a captura. Tais valores representam que a maioria dos inquiridos usufruiriam mais da ferramenta após do que durante a captura. Isto pode ser explicado pela falta de hábito existente dos utilizadores de anotar um vídeo enquanto está a ser capturado. O facto de anotar um vídeo já capturado traz indubitavelmente vantagens relativamente a anotar um vídeo que está a ser capturado, como o facto de existir acesso ao conteúdo total do vídeo e a possibilidade de rever incessantemente cada momento. A tarefa de anotação torna-se mais facilitada, precisa e completa, daí, a maior taxa de preferência de anotação ser associada a vídeos capturados.



**Figura 5.3:** Resultado da classificação de frequência do eventual uso dos cenários de utilização.

### 5.2.3.3 Tipos de Anotação

De seguida, foi pedido aos utilizadores que respondessem com que frequência inseririam cada tipo de anotação durante e após a captura. As respostas a esta pergunta estão compiladas no gráfico 5.3 e o seu objectivo é perceber quais os tipos de anotação geralmente mais adequados para os dois momentos. Chegou-se à conclusão que, conforme as expectativas, os tipos de anotação tinta digital e áudio obtiveram, em média, as maiores taxas de frequência (4,33 e 4,25 de 0 a 5) durante a captura, muito por causa da sua simplicidade e rapidez de inserção. Após a captura, os restantes tipos de anotação, com destaque para o tipo texto com média de 4,67, que tem um método de inserção mais demorado, obtiveram taxas de frequência altas.



**Figura 5.4:** Resultado da classificação de frequência do eventual uso dos tipos de anotação em cada cenário de utilização.

Em termos de facilidade de interagir com cada tipo de anotação, propriamente em marcar o seu início e fim, foram reunidos resultados positivos, onde, numa escala de 1 (difícil) a 5 (fácil) todos

os tipos tiveram classificações altas. O tipo de anotação áudio chegou mesmo a obter 100% no valor 5, ou seja, todos os inquiridos acharam a sua inserção fácil e natural, já que para isso basta pressionar o botão respectivo. As marcas e texto foram os tipos menos fáceis de utilizar, resultando, no entanto, numa média positiva. Inferiu-se que a mudança de posição das marcas e a inserção de anotações de texto apresentaram algumas dificuldades. Estas operações foram melhoradas, através da redução do número de fases para executar cada uma.

#### 5.2.3.4 Modos de Anotação

Quanto aos modos de anotação ilustrados no gráfico 5.4, evidenciou-se que os inquiridos usariam mais o modo atrasado e o modo suspenso do que o modo contínuo. Na escala de 1 (raramente) a 5 (frequentemente), em média o modo contínuo obteve apenas 2,83 contra os 4,50 e 4,33 do modo atrasado e suspenso respectivamente. Assim, houve uma escolha bem acentuada no modo atrasado como o modo principal de anotação, com recurso ao modo suspenso para anotar instantes do vídeo. Esta foi a metodologia de trabalho preferida na ferramenta entre os inquiridos e o coreógrafo Rui Horta, que a elogiou bastante. No que toca à usabilidade, em média, todos os modos foram bem percebidos e fáceis de usar.

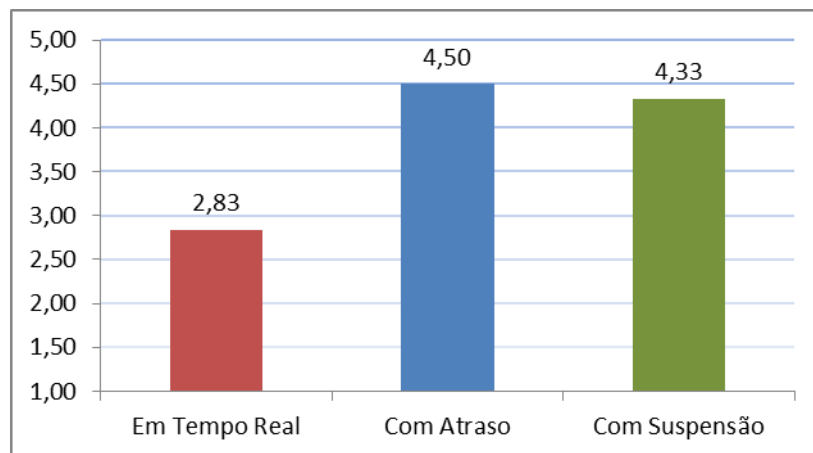


Figura 5.5: Resultado da classificação de frequência do eventual uso dos modos de anotação.

#### 5.2.3.5 Experiência

Em relação à envolvente emocional, a experiência caracterizou-se por ser atractiva (75%), útil (75%), cooperante (67%), clara (67%) e economizadora de tempo (67%). Com estes resultados, entende-se que há um interesse por parte dos artistas ligados à dança em possuírem um utensílio cooperante e útil no seu processo criativo, poupando assim tempo de trabalho. Esta caracterização podia ser revelada a partir da escolha de um conjunto de 48 expressões, baseadas no Microsoft Product Reaction Cards, que melhor definiam a experiência de utilização da solução implementada. Das 48, 29 foram

escolhidas, sendo que nesse sub conjunto escolhido, apenas 1 expressão negativa foi seleccionada: “muito técnica” com 8%, que corresponde a 1 utilizador. A escolha desta opção pode eventualmente dever-se ao facto de o utilizador não se sentir familiar e, por isso, confortável com as tecnologias aplicadas na solução.

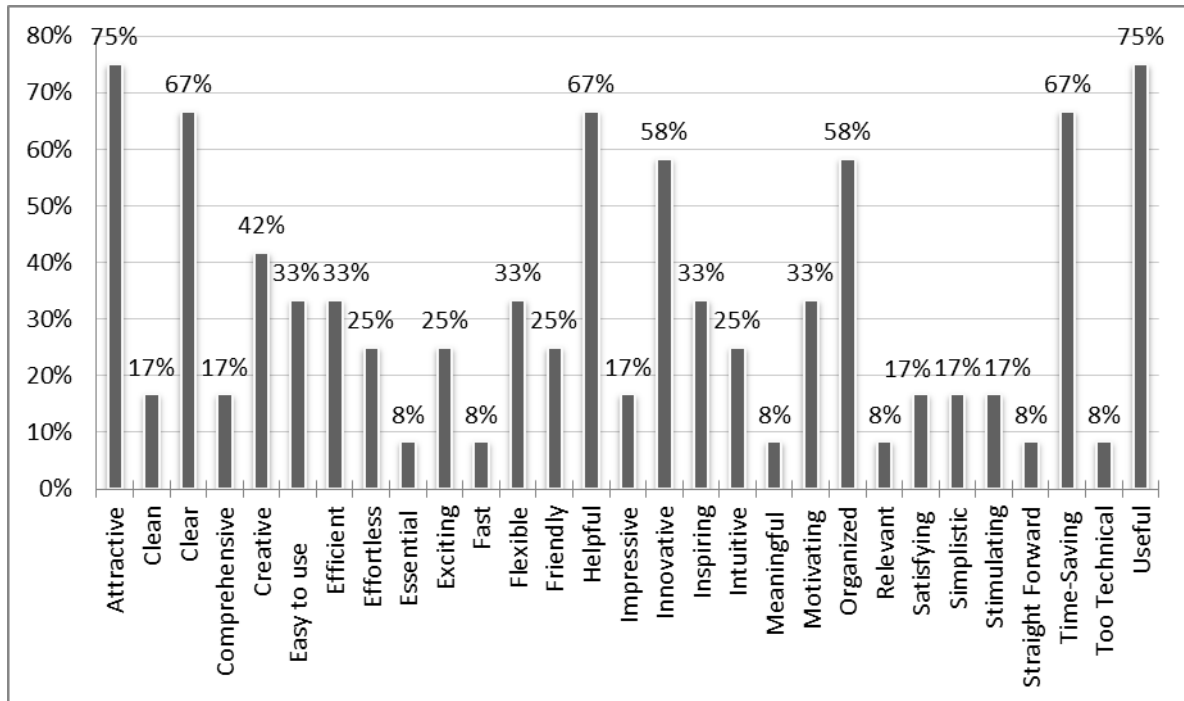


Figura 5.6: Resultado da classificação emocional da experiência.



## **6. Conclusões e Trabalho Futuro**

Neste capítulo é apresentada uma análise do trabalho realizado no âmbito desta dissertação, assim como sugestões e direcções a seguir relativamente à evolução da solução apresentada.

### **6.1 Conclusões**

Esta dissertação apresenta um anotador multimodal de vídeo em tempo real, usado como uma ferramenta de criação para a dança contemporânea. O protótipo, desenvolvido para Tablet PCs, explora a interacção com caneta digital e toque. Permite a associação de anotações durante a captura do vídeo em forma de tinta digital, áudio, texto, hiperligações e marcas. Suplementarmente, foi implementada uma interface para ser executada num dispositivo móvel, para controlar remotamente o sistema.

A ferramenta foi concebida para suportar o processo criativo dos coreógrafos, ao oferecer assistência para registar trabalhos durante os ensaios, na forma de vídeo com anotações associadas. No final das peças, pode ser complementado com anotações mais detalhadas, originando a construção de um arquivo individual de dança contemporânea. Por fazer parte do projecto TKB, este protótipo combina a experiência dos criadores contemporâneos, a fim de chegar a uma ferramenta original para composição, anotação e documentação de performance, onde os princípios conceptuais de cada coreógrafo são palavras-chave para a indexação e estrutura do seu arquivo. O sistema também pode ser aplicado em outras áreas onde a anotação da performance do corpo humano seja importante, como educação, teatro ou desportos de alta competição.

A avaliação seguida neste trabalho foi um processo regular ao longo do desenvolvimento, o que provocou frequentes iterações de desenho e implementação com base em melhoramentos a nível funcional. Contou com a presença de alunos e coreógrafos de dança, tais como Rui Horta e Stephan Jürgens, que inicialmente sugeriram novas funcionalidades, de modo a que a ferramenta satisfizesse as suas necessidades e expectativas e se tornasse um utensílio mais próximo da realidade e prático de usar. Posteriormente, foi realizada a avaliação da usabilidade do sistema via inquérito e observação da

interacção. Os resultados, em geral, foram positivos e constatou-se que a utilidade e a atractividade foram os factores que mais caracterizaram a experiência com a ferramenta. Verificou-se também que houve uma óbvia predilecção nos tipos de anotação tinta digital e áudio, com atraso e potencial suspensão na visualização da captura do vídeo.

Além do trabalho desenvolvido nesta dissertação, foram escritos e publicados dois artigos científicos sobre o sistema em duas conferências internacionais. O primeiro foi publicado na ACM CHI 2011 WIP [28], a conferência mais conceituada de interacção pessoa-máquina e o segundo foi publicado como *short paper* na ACM Multimedia 2011, a principal conferência mundial de multimédia.

## 6.2 Trabalho Futuro

No estado actual, a ferramenta é orientada a propósitos individuais, no entanto, seria mais interessante se houvesse um suporte para partilha e exportação da informação gerada. Como trabalho futuro, irá ser implementado e integrado um arquivo geral baseado na web e, assim, permitir este contacto com o exterior, tornando-se uma ferramenta mais colectiva.

Um método de reconhecimento de voz também poderá ser aplicado no futuro, para transcrever as anotações de voz para texto. Deste modo, podem ser indexadas, procuradas e recuperadas quando houver necessidade.

A versão *dark* da interface encontra-se num estado anterior ao estado da versão *light* e, por isso, terá de ser ajustada e concluída. Tal acontece, já que a versão *light* foi a versão escolhida para implementar e avaliar as novas mudanças no final do desenvolvimento. Com isto, a versão *dark* ficou num estado desactualizado.

Uma integração mais completa com a funcionalidade de *motion tracking* e o desenvolvimento de uma versão *Mac* são outras tarefas programadas para o futuro.

Acreditamos que este sistema contribuiu fortemente para a área em comparação com os trabalhos relacionados, e tal justifica-se pela aceitação de dois artigos científicos em conferências com taxas de aprovação reduzidas e pelo grande número de comentários positivos recebidos dos utilizadores alvo. Observou-se que, quando aplicada à dança contemporânea, estas anotações criadas em tempo real associadas ao vídeo conseguem criar uma nova forma de comunicação mais efectiva e directa entre o bailarino e o coreógrafo, aumentando em várias formas o tipo de *feedback*, para realizar o derradeiro objectivo num ensaio, que é a busca pela perfeição na performance do artista.



## Bibliografia

- [1] R. Goularte, *et al.*, "M4Note: A Multimodal Tool for Multimedia Annotations," presented at the Proceedings of the WebMedia & LA-Web 2004 Joint Conference 10th Brazilian Symposium on Multimedia and the Web 2nd Latin American Web Congress, 2004.
- [2] R. Kannan, *et al.*, "DanVideo: an MPEG-7 authoring and retrieval system for dance videos," *Multimedia Tools Appl.*, vol. 46, pp. 545-572.
- [3] P. Wittenburg, *et al.*, "ELAN: a Professional Framework for Multimodality Research," presented at the Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation, 2006.
- [4] P. Bottoni, *et al.*, "MADCOW: a multimedia digital annotation system," presented at the Proceedings of the working conference on Advanced visual interfaces, Gallipoli, Italy, 2004.
- [5] D. C. A. Bulterman, "Animating Peer-Level Annotations Within Web-Based Multimedia," *Eurographics Multimedia*, pp. 49--57, 2004.
- [6] M. Shilman and W. E. I. Zile, *Recognizing freeform digital ink annotations* vol. 3163. Berlin, ALLEMAGNE: Springer, 2004.
- [7] N. Correia and T. Chambel, "Active video watching using annotation," presented at the Proceedings of the seventh ACM international conference on Multimedia (Part 2), Orlando, Florida, United States, 1999.
- [8] (2004, 05.12.2010). *MPEG-7 Overview (version 10)*. Available: <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>
- [9] M. J. Potel and R. E. Sayre, "Interacting with the GALATEA film analysis system," *SIGGRAPH Comput. Graph.*, vol. 10, pp. 52-59, 1976.
- [10] W. E. Mackay, "EVA: an experimental video annotator for symbolic analysis of video data," *SIGCHI Bull.*, vol. 21, pp. 68-71, 1989.
- [11] B. L. Harrison and R. M. Baecker, "Designing video annotation and analysis systems," presented at the Proceedings of the conference on Graphics interface '92, Vancouver, British Columbia, Canada, 1992.

- [12] C.-Y. Lin, *et al.* *Videoannex annotation tool*. Available: <http://www.research.ibm.com/VideoAnnEx/>
- [13] Bargeron, *et al.*, *Annotations for streaming video on the Web: system design and usage studies*. Kidlington, ROYAUME-UNI: Elsevier, 1999.
- [14] B. Ramadoss and K. Rajkumar, "Generic Modeling and Annotation of the Dance Video Semantics," presented at the Proceedings of the Sixth IEEE International Conference on Computer and Information Technology, 2006.
- [15] G. Cherry, *et al.* (2003). *Using a Video Annotation Tool to Teach Dance Composition*. Available: <http://imej.wfu.edu/articles/2003/1/01/printver.asp>
- [16] D. Cabral and N. Correia, "Pen-Based Video Annotations: A Proposal and a Prototype for Tablet PCs," presented at the Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II, Uppsala, Sweden, 2009.
- [17] K. Weher and A. Poon, "Marquee: a tool for real-time video logging," presented at the Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence, Boston, Massachusetts, United States, 1994.
- [18] G. Rudinei, *et al.*, "Structuring interactive TV documents," presented at the Proceedings of the 2003 ACM symposium on Document engineering, Grenoble, France, 2003.
- [19] P. Chiu, *et al.*, "NoteLook: taking notes in meetings with digital video and ink," presented at the Proceedings of the seventh ACM international conference on Multimedia (Part 1), Orlando, Florida, United States, 1999.
- [20] G. Ramos and R. Balakrishnan, "Fluid interaction techniques for the control and annotation of digital video," presented at the Proceedings of the 16th annual ACM symposium on User interface software and technology, Vancouver, Canada, 2003.
- [21] W3C. (2008). *Synchronized Multimedia Integration Language*. Available: <http://www.w3.org/TR/smil/>
- [22] N. Diakopoulos and I. Essa, "Videotater: an approach for pen-based digital video segmentation and tagging," presented at the Proceedings of the 19th annual ACM symposium on User interface software and technology, Montreux, Switzerland, 2006.
- [23] P. Brandl, *et al.*, "Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces," presented at the Proceedings of the working conference on Advanced visual interfaces, Napoli, Italy, 2008.
- [24] K. Hinckley, *et al.*, "Pen + touch = new tools," presented at the Proceedings of the 23rd annual ACM symposium on User interface software and technology, New York, New York, USA.
- [25] "iPhone Human Interface Guidelines for Web Applications: User Experience," Apple Inc., Technical Report2010.

- [26] (2002, 20.06.2011). *The Open Sound Control 1.0 Specification*. Available: [http://opensoundcontrol.org/spec-1\\_0](http://opensoundcontrol.org/spec-1_0)
- [27] R. Dunlop. *Catmull-Rom Splines*. Available: <http://www.mvps.org/directx/articles/catmull/>
- [28] D. Cabral, *et al.*, "Multimodal video annotation for contemporary dance creation," presented at the Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems, Vancouver, BC, Canada, 2011.



## **Anexo A – XML Schema**

Como referido ao longo do documento apresentado, a tecnologia XML foi utilizada na aplicação desenvolvida com o intuito de recolher e organizar de forma estruturada todas as informações semânticas associadas ao processo de anotação. Posto isto, o conjunto de diagramas que se segue apresentam o XML Schema correspondente a extractos do documento XML. Cada documento armazena as marcas criadas e as anotações associadas a um vídeo.

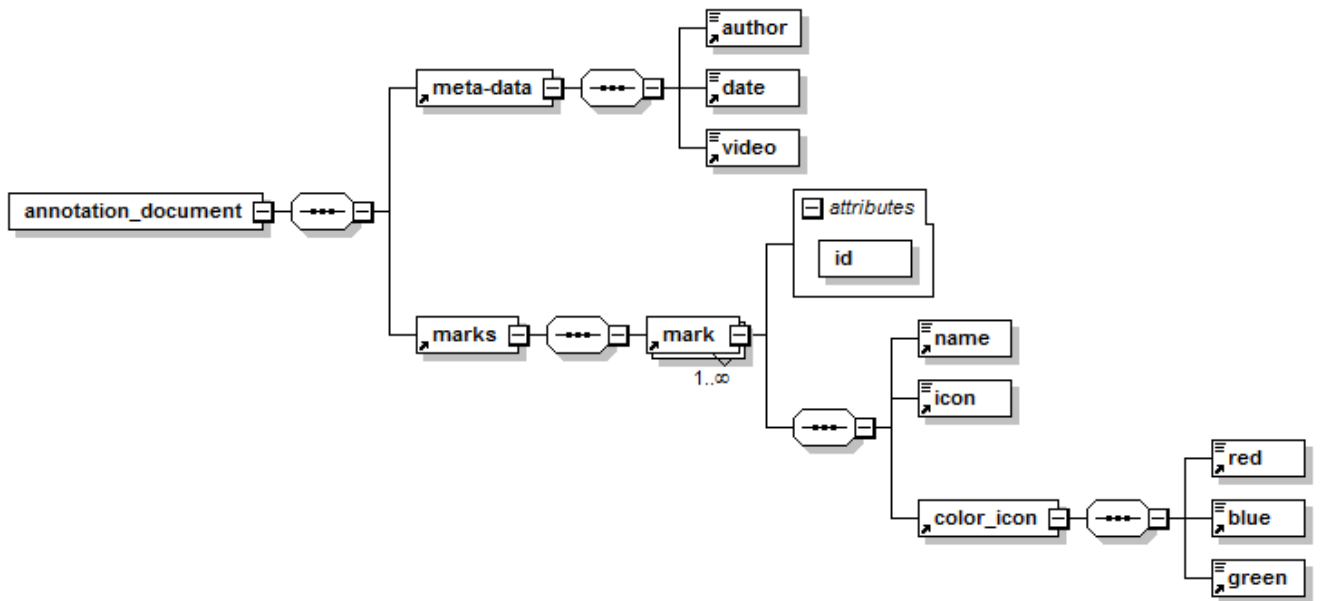


Figura A.1: XML Schema para o armazenamento das marcas de sistema e personalizadas existentes para aplicação.

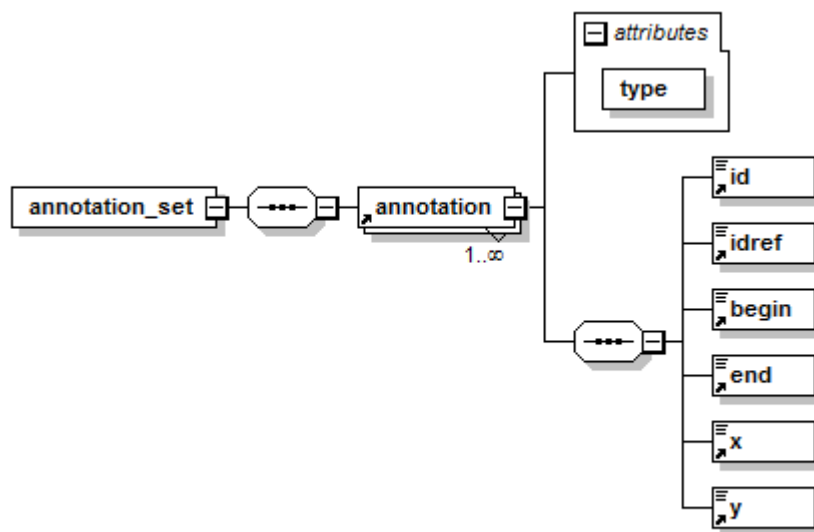


Figura A.2: XML Schema para armazenamento das anotações do tipo marcas.

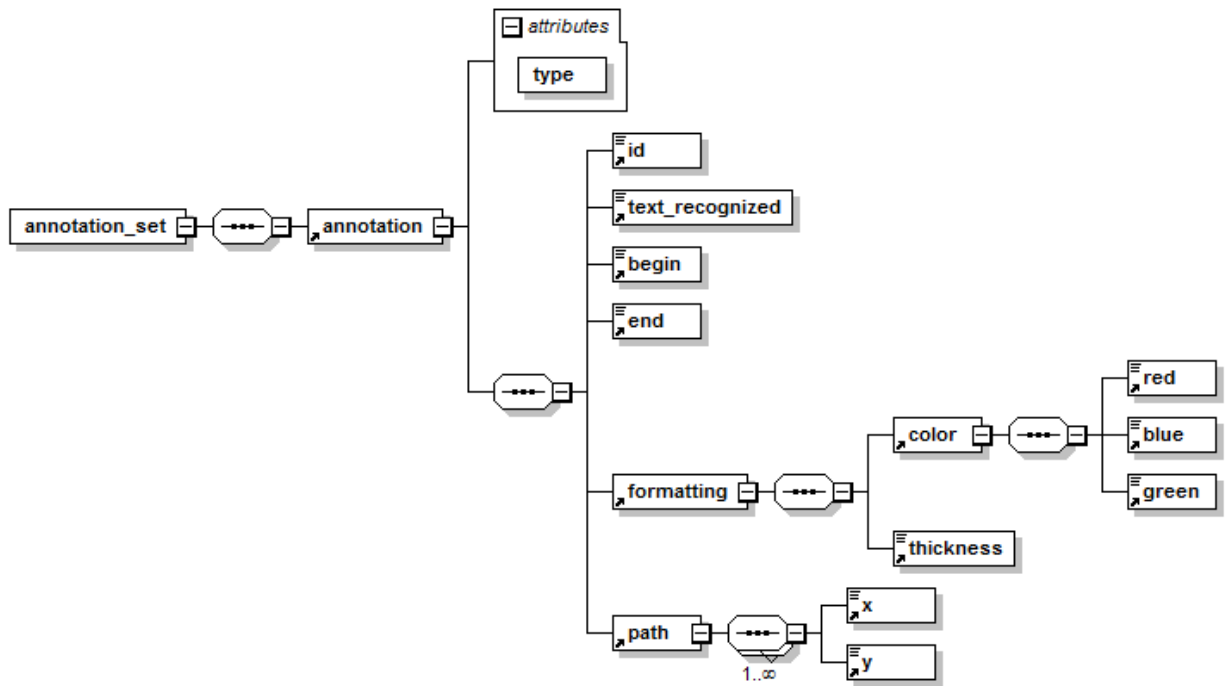


Figura A.3: XML Schema para armazenamento das anotações do tipo tinta digital.

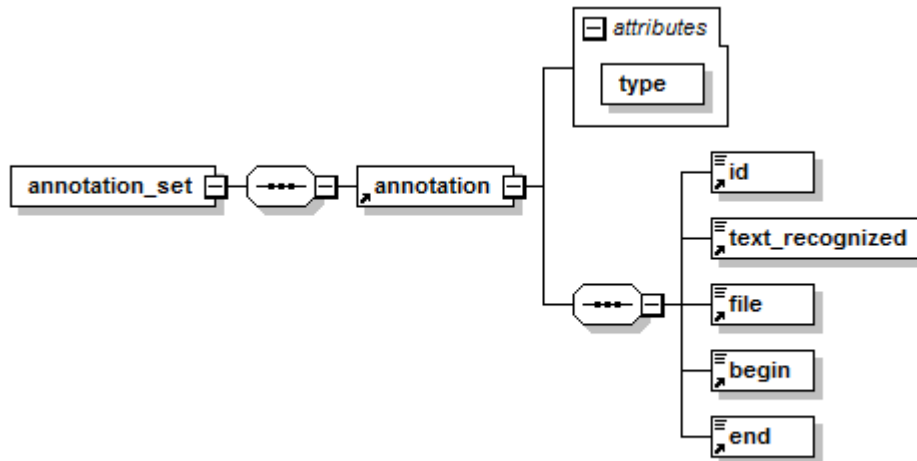


Figura A.4: XML Schema para armazenamento das anotações do tipo áudio.

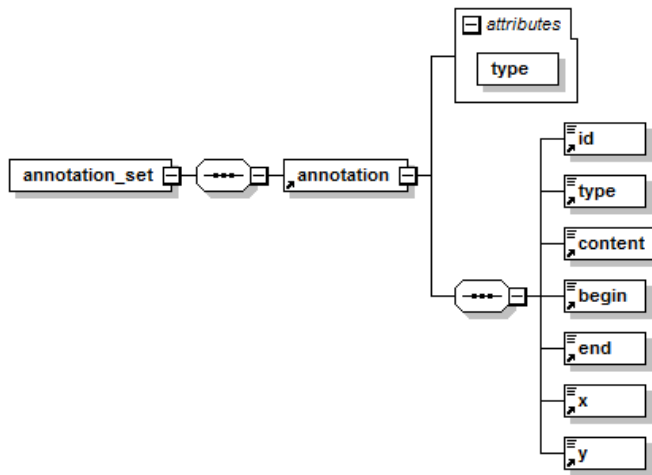


Figura A.5: XML Schema para armazenamento das anotações do tipo hiperligações.

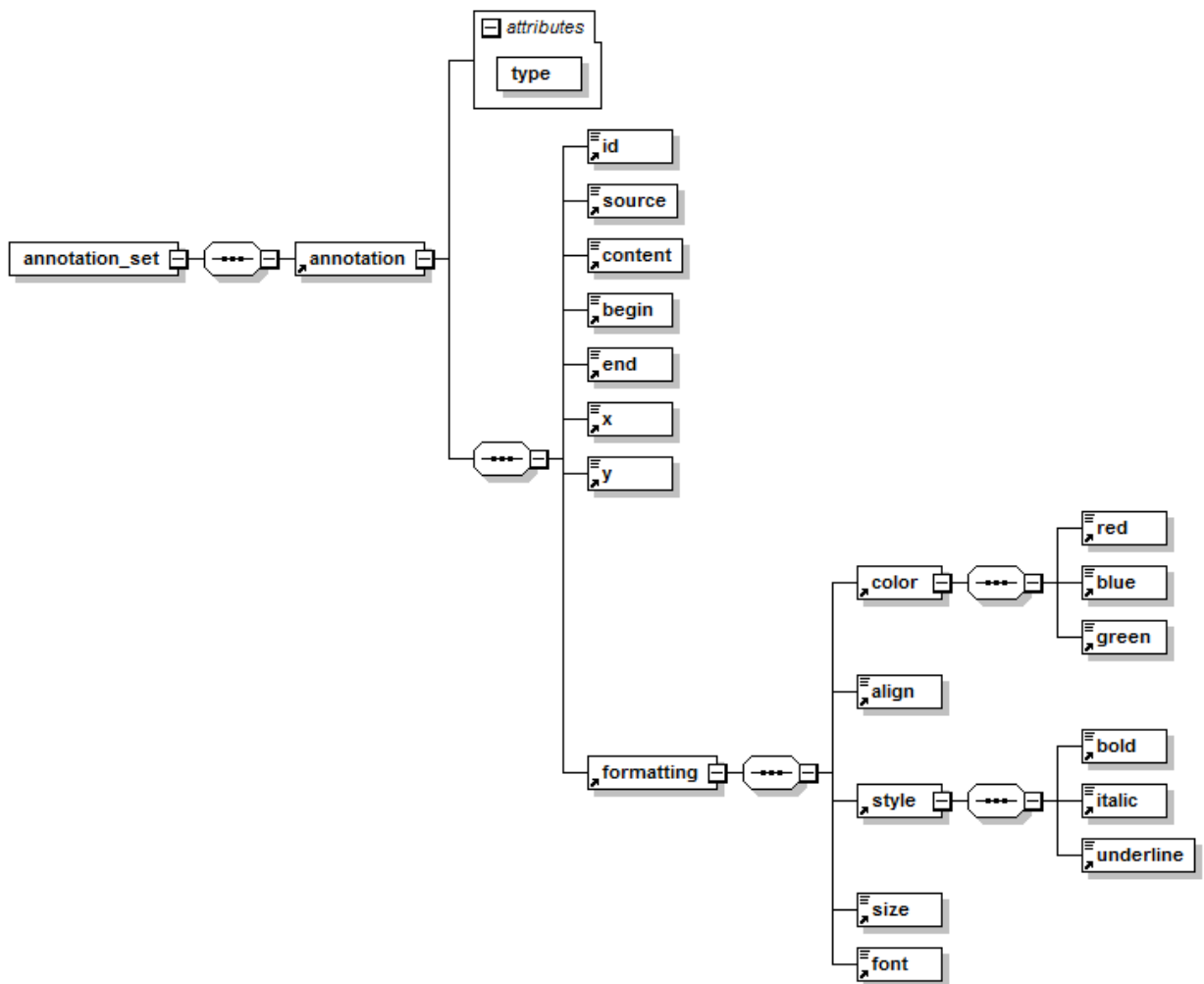


Figura A.6: XML Schema para armazenamento das anotações do tipo texto.



## **Anexo B – Questionário**

O modelo completo do questionário utilizado na avaliação através de inquérito aos utilizadores é apresentado de seguida.

# Enunciado

## Questionnaire – Creation-Tool

1. Consider two types of scenarios, during a rehearsal (in video capture) and after a rehearsal (after video capture). Please, compare and classify how often are you willing to use the Creation-Tool, in each scenario:

	Rarely				Frequently
During a rehearsal	1	2	3	4	5
After a rehearsal	1	2	3	4	5

2. Consider the different annotation types of the Creation-Tool. Please, compare and classify how often are you willing to use each type, during a rehearsal.

	Rarely				Frequently
Sketch	1	2	3	4	5
Text	1	2	3	4	5
Audio	1	2	3	4	5
Marks	1	2	3	4	5
Hyperlinks	1	2	3	4	5

3. Consider the different annotation types of the Creation-Tool. Please, compare and classify how often are you willing to use each type, after a rehearsal.

	Rarely				Frequently
Sketch	1	2	3	4	5
Text	1	2	3	4	5
Audio	1	2	3	4	5
Marks	1	2	3	4	5
Hyperlinks	1	2	3	4	5

4. Consider the insertion of different annotation types of the Creation-Tool. Please, compare and classify them.

	Difficult				Easy
Sketch	1	2	3	4	5
Text	1	2	3	4	5
Audio	1	2	3	4	5
Marks	1	2	3	4	5
Hyperlinks	1	2	3	4	5

5. Consider the continuous and suspended annotation modes of the Creation-Tool. Please, compare and classify how often are you willing to use each mode.

	Rarely				Frequently
Continuous Mode	1	2	3	4	5
Suspended Mode	1	2	3	4	5

6. Consider the continuous and suspended annotation modes of the Creation-Tool. Please, compare and classify them.

	Difficult				Easy
Continuous Mode	1	2	3	4	5
Suspended Mode	1	2	3	4	5

7. Consider the real-time and delayed video modes of the Creation-Tool. Please, compare and classify how often are you willing to use each mode, during a rehearsal.

	Rarely				Frequently
Real-Time Mode	1	2	3	4	5
Delayed Mode	1	2	3	4	5

8. Consider your interaction with the Creation-Tool. Please, classify it.

Difficult				Easy	
1	2	3	4	5	

**9. Choose the expression(s) that better classifies your experience with the Creation-Tool.**

- |  |   |                                     |   |
|--|---|-------------------------------------|---|
| <input type="checkbox"/> Annoying      | <input type="checkbox"/> Essential        | <input type="checkbox"/> Innovative | <input type="checkbox"/> Simplistic       |
| <input type="checkbox"/> Attractive    | <input type="checkbox"/> Exceptional      | <input type="checkbox"/> Inspiring  | <input type="checkbox"/> Slow             |
| <input type="checkbox"/> Clean         | <input type="checkbox"/> Exciting         | <input type="checkbox"/> Intuitive  | <input type="checkbox"/> Sterile          |
| <input type="checkbox"/> Clear         | <input type="checkbox"/> Familiar         | <input type="checkbox"/> Irrelevant | <input type="checkbox"/> Stimulating      |
| <input type="checkbox"/> Complex       | <input type="checkbox"/> Fast             | <input type="checkbox"/> Meaningful | <input type="checkbox"/> Straight Forward |
| <input type="checkbox"/> Comprehensive | <input type="checkbox"/> Flexible         | <input type="checkbox"/> Motivating | <input type="checkbox"/> Stressful        |
| <input type="checkbox"/> Confusing     | <input type="checkbox"/> Friendly         | <input type="checkbox"/> Novel      | <input type="checkbox"/> Time-consuming   |
| <input type="checkbox"/> Creative      | <input type="checkbox"/> Frustrating      | <input type="checkbox"/> Old        | <input type="checkbox"/> Time-Saving      |
| <input type="checkbox"/> Distracting   | <input type="checkbox"/> Hard to Use      | <input type="checkbox"/> Organized  | <input type="checkbox"/> Too Technical    |
| <input type="checkbox"/> Easy to use   | <input type="checkbox"/> Helpful          | <input type="checkbox"/> Powerful   | <input type="checkbox"/> Unattractive     |
| <input type="checkbox"/> Efficient     | <input type="checkbox"/> Impressive       | <input type="checkbox"/> Relevant   | <input type="checkbox"/> Unconventional   |
| <input type="checkbox"/> Effortless    | <input type="checkbox"/> Incomprehensible | <input type="checkbox"/> Satisfying | <input type="checkbox"/> Useful           |

**10. Comments and Suggestions:**

---

---

---

---

---

---

---

---

11. Do you usually record your work?  Yes  No

If yes, in which media?  Paper Notebook  
 Audio  
 Video  
 Other: \_\_\_\_\_

12. Do you usually annotate during your work process?  Yes  No

If yes, in which technology?  Paper Notebook  
 Desktop Computer  
 Laptop Computer  
 Mobile Phone  
 PDA  
 Tablet  
 Other: \_\_\_\_\_

13. Do you usually share your work documents?  Yes  No

If yes, in which platform?  Mail (Post)  
 Fax  
 Hardware (Sharing CDs, DVDs, Pen Drives, etc..)  
 E-mail  
 Web Site (personal site, blog, etc..)  
 Instant Messaging (Skype, MSN, etc..)  
 Social Network (Facebook, Myspace, etc..)  
 Other: \_\_\_\_\_

14. Were you familiar with pen-based technology (Tablets or PDAs) before:  Yes  No

15. Were you familiar with touch-based technology (Tablets or Touch Phones) before:  Yes  No

16. Age: \_\_\_\_\_

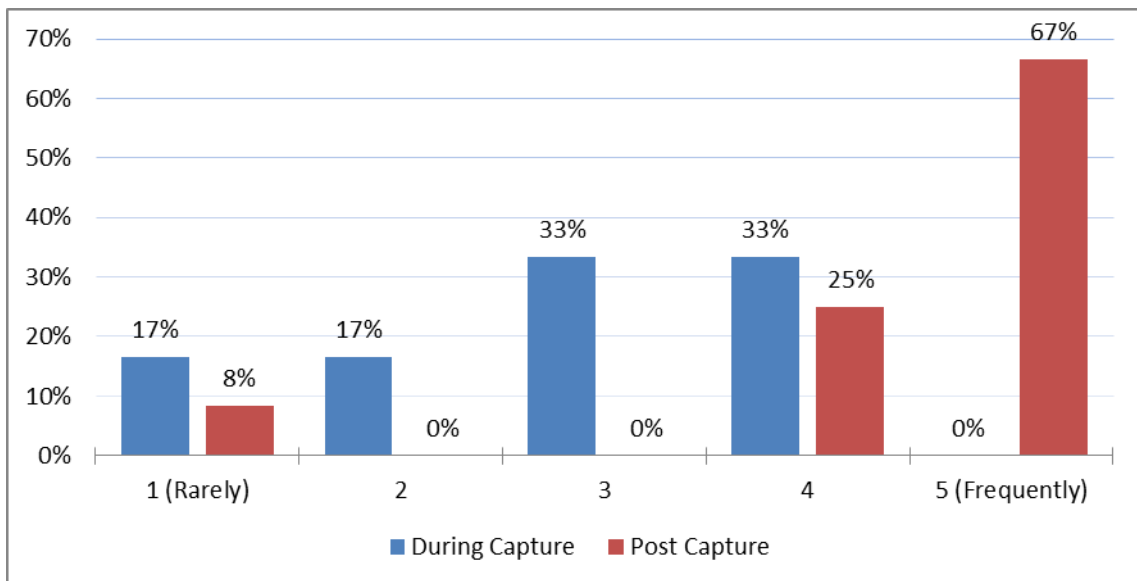
17. Genre:  Masculine  Feminine

18. Education:  Elementary School  High School  Bachelor  Master  PhD

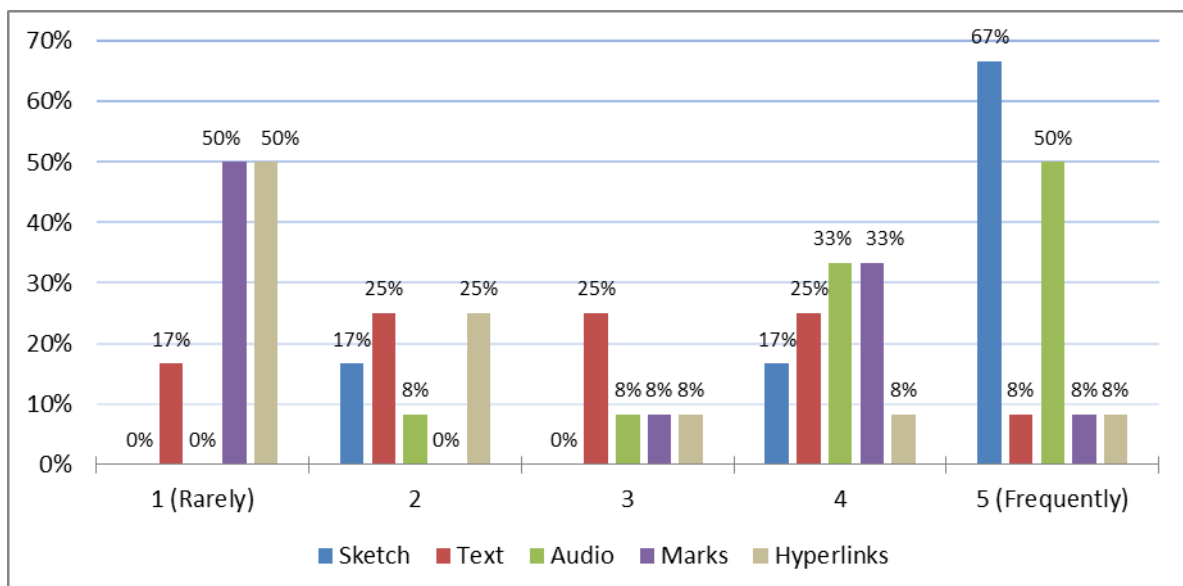
19. Handedness:  Left  Right

# Resultados

1. Consider two types of scenarios, during a rehearsal (in video capture) and after a rehearsal (after video capture). Please, compare and classify how often are you willing to use the Creation-Tool, in each scenario:

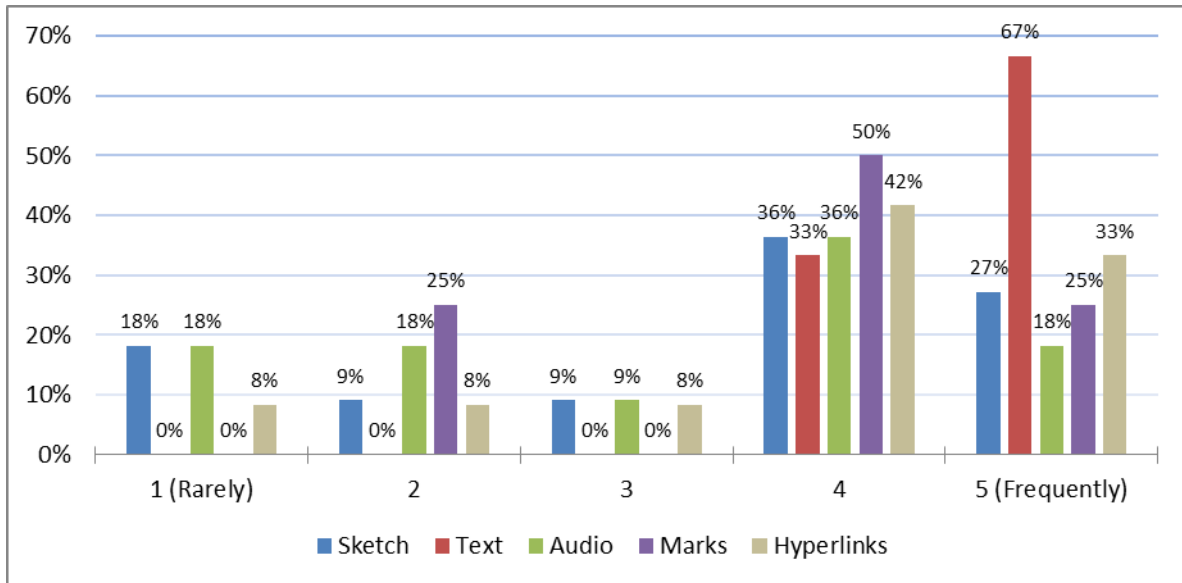


2. Consider the different annotation types of the Creation-Tool. Please, compare and classify how often are you willing to use each type, during a rehearsal.

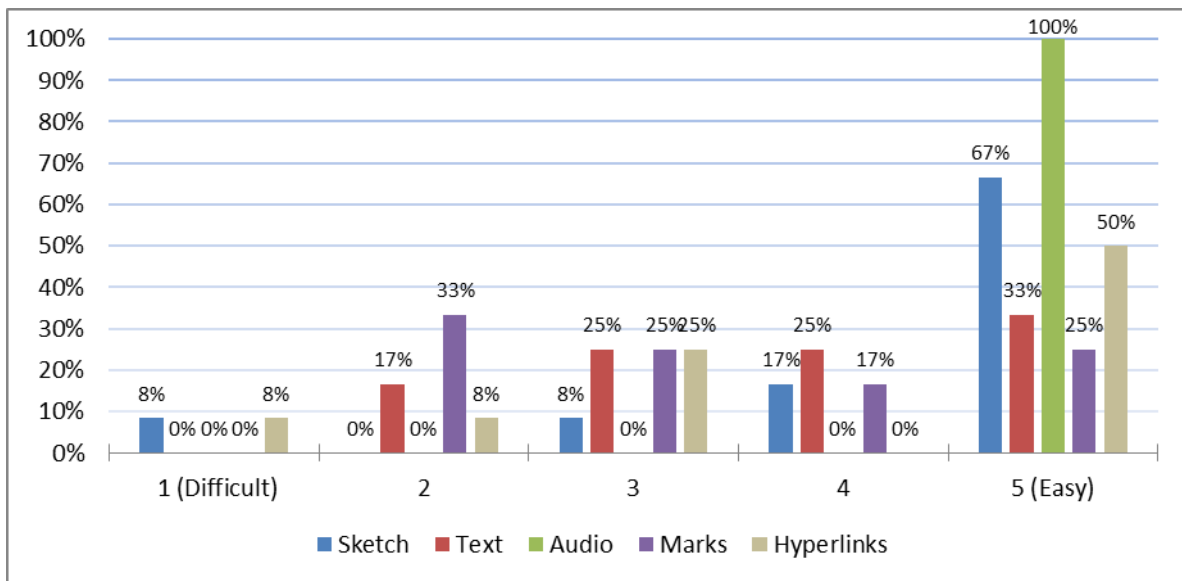




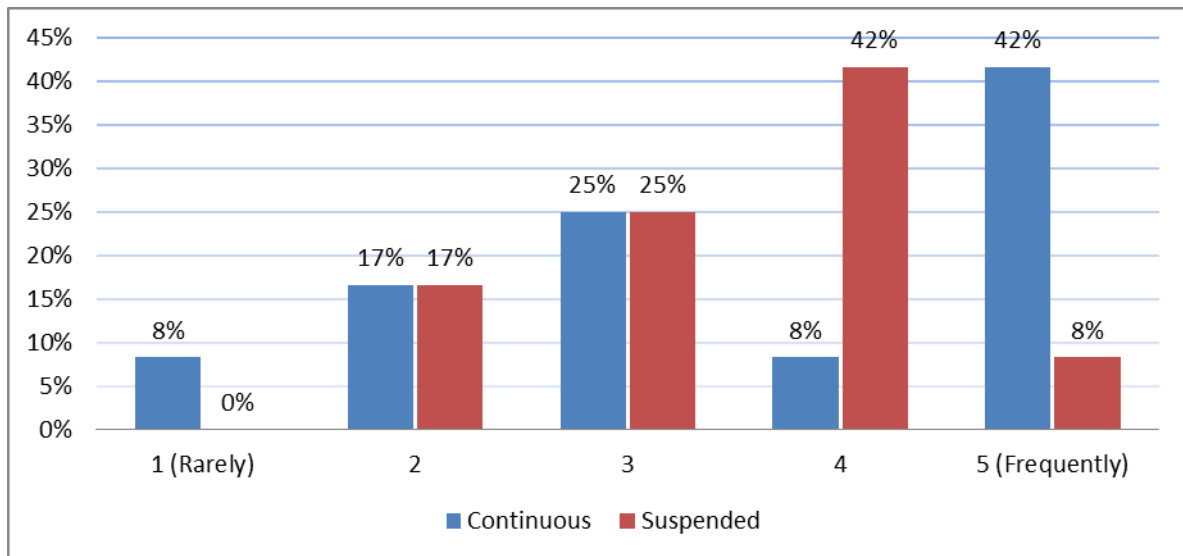
3. Consider the different annotation types of the Creation-Tool. Please, compare and classify how often are you willing to use each type, after a rehearsal.



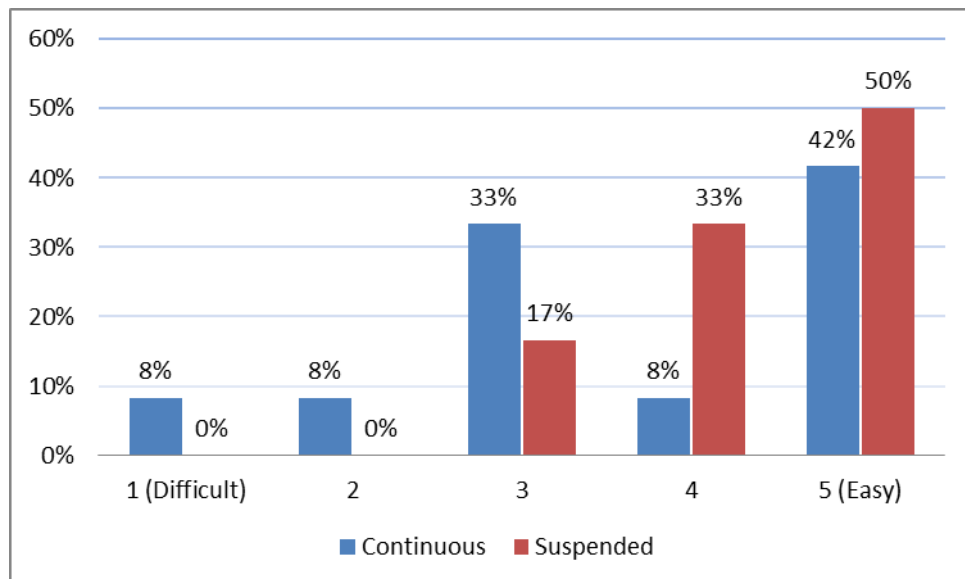
4. Consider the insertion of different annotation types of the Creation-Tool. Please, compare and classify them.



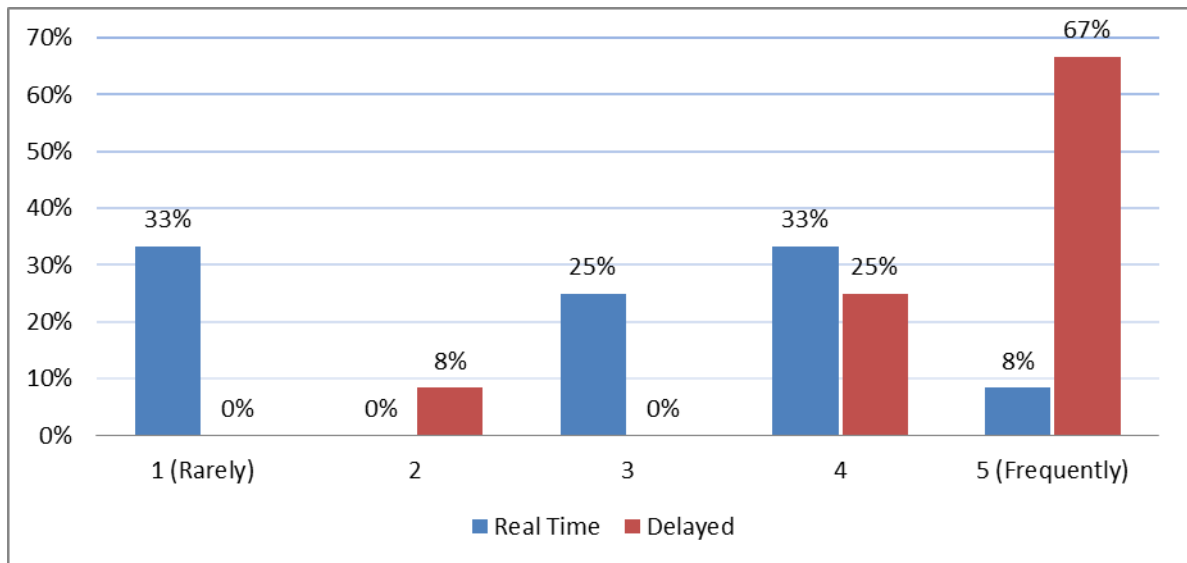
5. Consider the continuous and suspended annotation modes of the Creation-Tool. Please, compare and classify how often are you willing to use each mode.



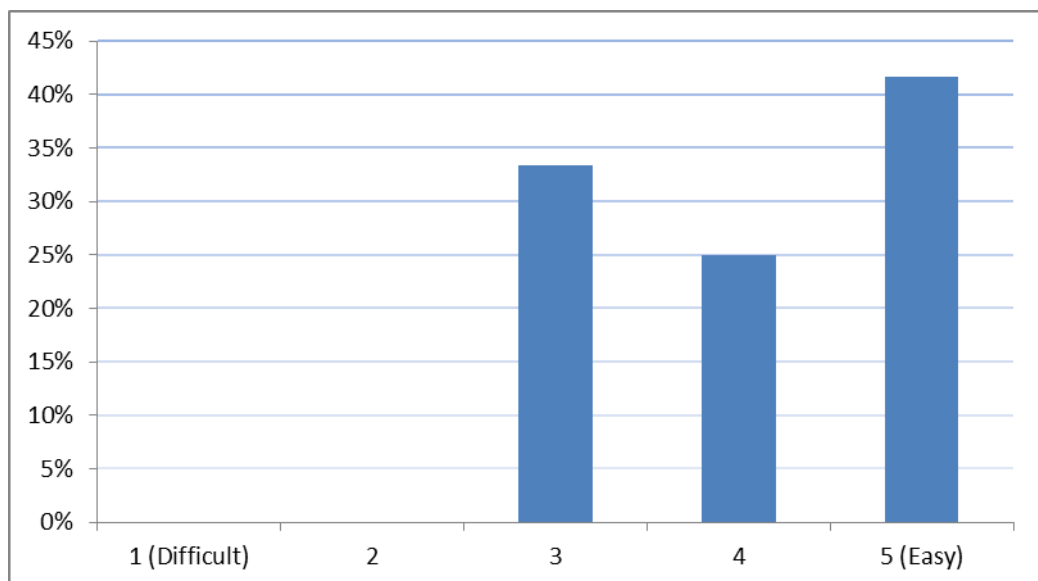
6. Consider the continuous and suspended annotation modes of the Creation-Tool. Please, compare and classify them.



7. Consider the real-time and delayed video modes of the Creation-Tool. Please, compare and classify how often are you willing to use each mode, during a rehearsal.

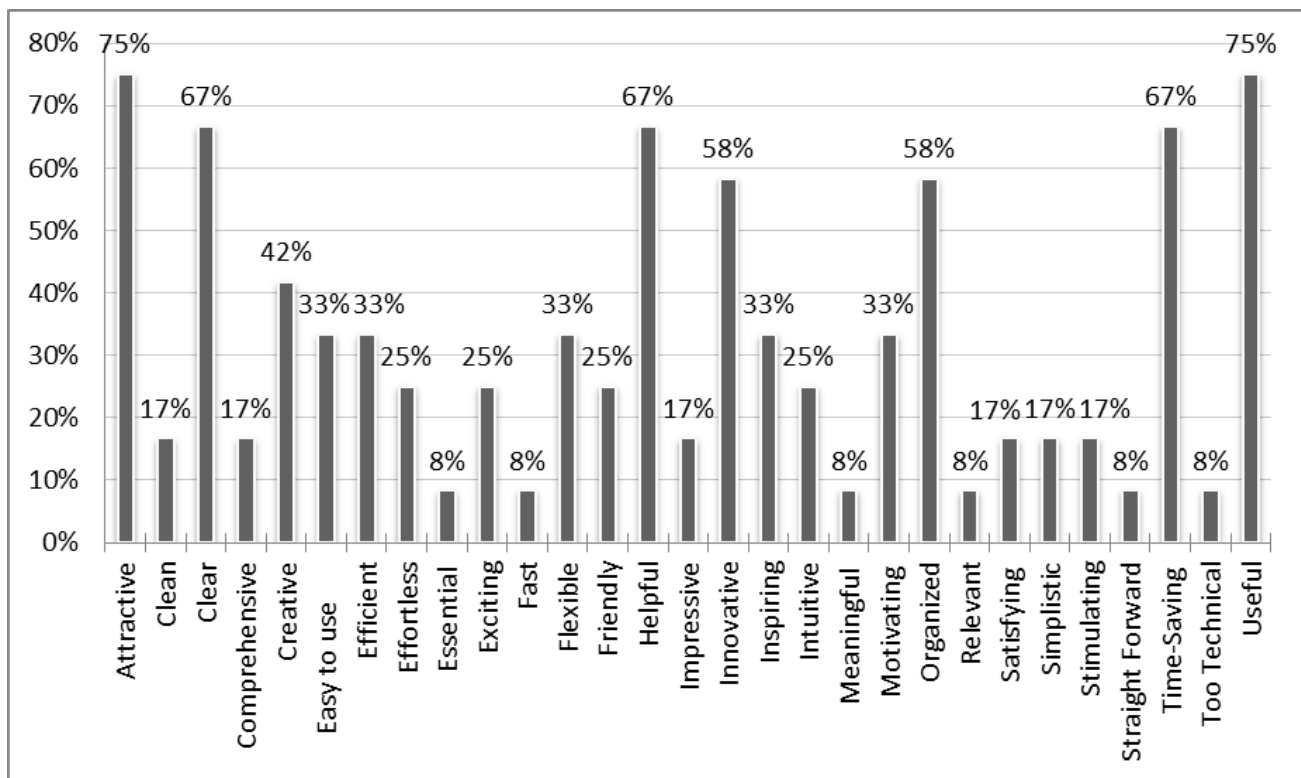


8. Consider your interaction with the Creation-Tool. Please, classify it.

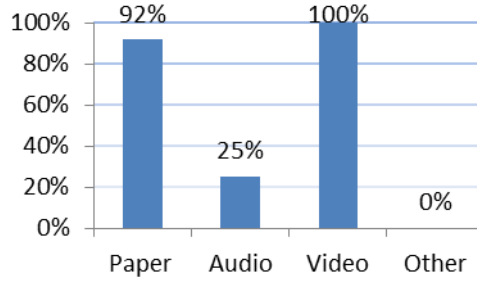
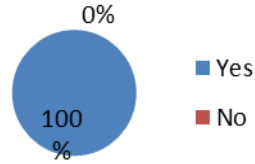


**9. Choose the expression(s) that better classifies your experience with the Creation-Tool.**

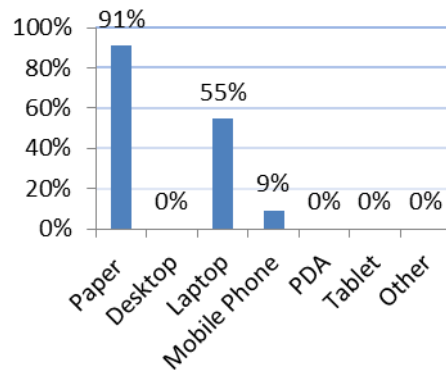
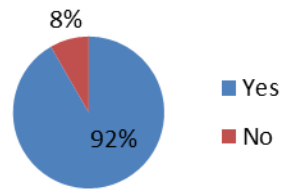
- |  |   |                                     |   |
|--|---|-------------------------------------|---|
| <input type="checkbox"/> Annoying      | <input type="checkbox"/> Essential        | <input type="checkbox"/> Innovative | <input type="checkbox"/> Simplistic       |
| <input type="checkbox"/> Attractive    | <input type="checkbox"/> Exceptional      | <input type="checkbox"/> Inspiring  | <input type="checkbox"/> Slow             |
| <input type="checkbox"/> Clean         | <input type="checkbox"/> Exciting         | <input type="checkbox"/> Intuitive  | <input type="checkbox"/> Sterile          |
| <input type="checkbox"/> Clear         | <input type="checkbox"/> Familiar         | <input type="checkbox"/> Irrelevant | <input type="checkbox"/> Stimulating      |
| <input type="checkbox"/> Complex       | <input type="checkbox"/> Fast             | <input type="checkbox"/> Meaningful | <input type="checkbox"/> Straight Forward |
| <input type="checkbox"/> Comprehensive | <input type="checkbox"/> Flexible         | <input type="checkbox"/> Motivating | <input type="checkbox"/> Stressful        |
| <input type="checkbox"/> Confusing     | <input type="checkbox"/> Friendly         | <input type="checkbox"/> Novel      | <input type="checkbox"/> Time-consuming   |
| <input type="checkbox"/> Creative      | <input type="checkbox"/> Frustrating      | <input type="checkbox"/> Old        | <input type="checkbox"/> Time-Saving      |
| <input type="checkbox"/> Distracting   | <input type="checkbox"/> Hard to Use      | <input type="checkbox"/> Organized  | <input type="checkbox"/> Too Technical    |
| <input type="checkbox"/> Easy to use   | <input type="checkbox"/> Helpful          | <input type="checkbox"/> Powerful   | <input type="checkbox"/> Unattractive     |
| <input type="checkbox"/> Efficient     | <input type="checkbox"/> Impressive       | <input type="checkbox"/> Relevant   | <input type="checkbox"/> Unconventional   |
| <input type="checkbox"/> Effortless    | <input type="checkbox"/> Incomprehensible | <input type="checkbox"/> Satisfying | <input type="checkbox"/> Useful           |



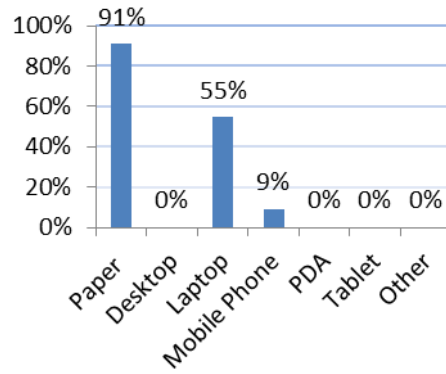
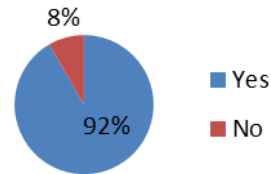
11. Do you usually record your work?



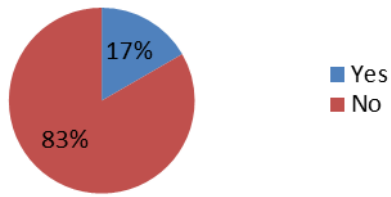
12. Do you usually annotate during your work process?



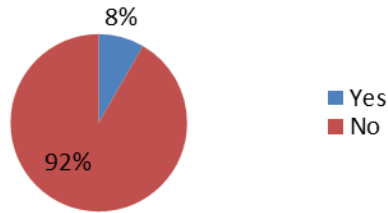
13. Do you usually share your work documents?



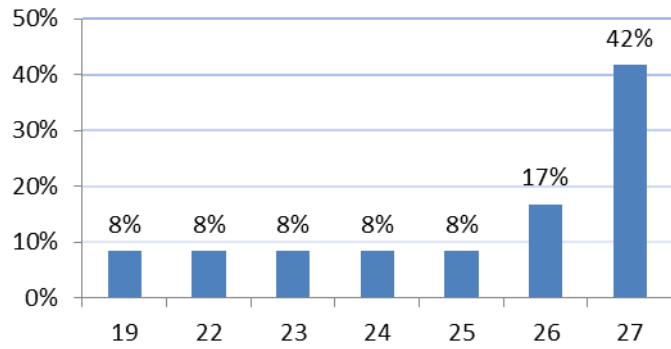
14. Were you familiar with pen-based technology (Tablets or PDAs) before:



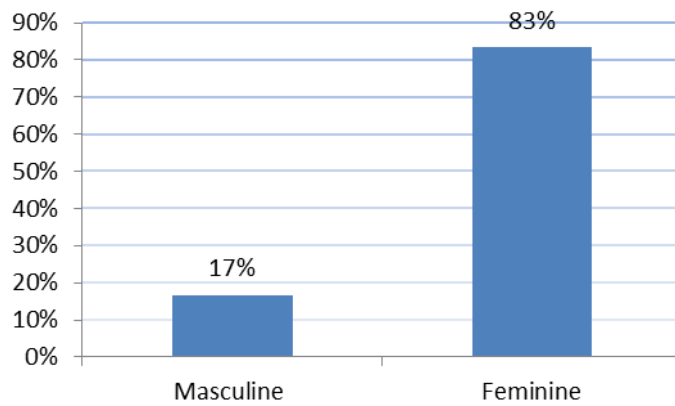
15. Were you familiar with touch-based technology (Tablets or Touch Phones) before:



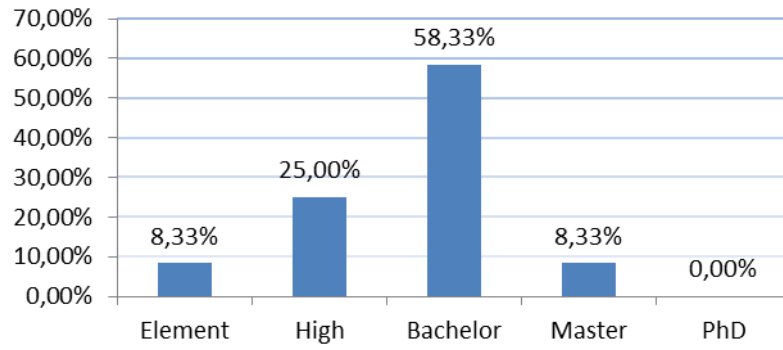
16. Age



17. Genre



**18. Education**



**19. Handedness:**

