

**Faculdade de Engenharia da Universidade do Porto**



**FEUP**

**Correspondência entre alterações da voz e do  
electrocardiograma**

Ricardo Jorge Macedo Pereira de Sousa Dias

ESTADO DA ARTE

Relatório de Preparação da Dissertação realizado no âmbito do  
Mestrado Integrado em Engenharia Eletrotécnica e de Computadores  
Major Telecomunicações, Eletrónica e Computadores

Orientador: Prof. Dr. Aníbal João de Sousa Ferreira

17 de fevereiro de 2012



# Índice

Índice.....	iii
Lista de figuras .....	iv
Abreviaturas e Símbolos .....	v
<b>Capítulo 1 .....</b>	<b>1</b>
Introdução.....	1
1.1 - Descrição do Tema .....	1
1.2 - Motivação .....	1
1.3 - Objetivos a atingir .....	2
1.4 - Estrutura .....	2
<b>Capítulo 2 .....</b>	<b>3</b>
Estado da Arte .....	3
2.1 - Caracterização do eletrocardiograma.....	3
2.1.1 - Enquadramento Fisiológico .....	3
2.1.2 - Aplicabilidade .....	4
2.1.3 - Eventos .....	5
2.1.4 - Principais riscos .....	6
2.1.5 - Dispositivos utilizados.....	7
2.1.6 - Exame típico.....	7
2.1.7 - Exemplos de anomalias .....	8
2.2 - Caracterização de parâmetros do som.....	8
2.2.1 - Intensidade sonora .....	8
2.2.2 - Frequência fundamental (F0) e formantes (F1, F2...Fn) .....	9
2.2.3 - Tonalidade, timbre e intensidade subjetiva .....	9
2.2.4 - Representação e análise de Fourier .....	9
2.2.5 - Espetro e espectrograma .....	9
2.2.6 - Tonalidade, timbre e intensidade subjetiva .....	9
2.3 - Produção e percepção da fala .....	10
2.3.1 - Percepção auditiva.....	10
2.3.2 - Produção de fala.....	11
2.3.3 - Modos e lugares de articulação .....	11
2.3.4 - Mecanismos de fonação.....	11
2.3.5 - Modelo fonte-filtro .....	12
2.4 - Medição de parâmetros de voz .....	12
2.4.1 - Critérios de avaliação.....	12
2.4.2 - Jitter .....	13
2.4.3 - Shimmer .....	13
2.4.4 - Harmonic-to-noise ratio (HNR) .....	13

2.4.5 - Tempo máximo de fonação .....	14
2.4.6 - Estimação de impulsos glotais (EGG).....	14
2.5 - Avanços na avaliação de alterações na voz e no ECG .....	14
2.5.1 - Modelos existentes .....	14
2.6 - Conclusões .....	15
<b>Capítulo 3 .....</b>	<b>16</b>
Ferramentas a utilizar .....	16
3.1 - Algoritmia de Processamento de Sinal .....	16
3.2 - Praat .....	16
3.3 - SEEGNAL Voice Studio.....	17
3.4 - Biopac MP36RWSW.....	17
<b>Capítulo 4 .....</b>	<b>18</b>
Estruturação do projeto .....	18
4.1 - Resultados esperados .....	18
4.2 - Principais desafios.....	18
4.3 - Plano de trabalho.....	19
<b>Capítulo 5 .....</b>	<b>20</b>
Conclusão .....	20
<b>Referências Bibliográficas.....</b>	<b>21</b>
<b>Referências Sitiográficas .....</b>	<b>22</b>

## Lista de figuras

<b>Figura 2.1</b> - Mecanismo de despolarização e repolarização .....	4
<b>Figura 2.2</b> - Alguns eventos de um ECG [1.1] .....	5
<b>Figura 2.3</b> - Principais dispositivos de um ECG [1.1] .....	7
<b>Figura 2.4</b> - Espetrograma de um violino [2.2] .....	10
<b>Figura 2.5</b> - Anatomia do ouvido humano [1.1] .....	11
<b>Figura 2.6</b> - Representação de impulsos glotais (à esquerda) e fórmula de cálculo do RAP (à direita) .....	13
<b>Figura 2.7</b> - Representação de impulsos glotais (à esquerda) e fórmula de cálculo do APQ (à direita) .....	13
<b>Figura 2.8</b> - Fórmula para o cálculo do HNR .....	14
<b>Figura 3.1</b> - Aparelho Biopac [2.3] .....	17
<b>Figura 4.1</b> - Diagrama de Gantt do projeto.....	19

# Abreviaturas e Símbolos

Lista de abreviaturas (ordenadas por ordem alfabética)

APQ	Amplitude Perturbation Quotient
DFT	Discrete Fourier Transform
ECG	Eletrocardiograma
HNR	Harmonic-to-noise ratio
FFT	Fast Fourier Transform
PPQ	Period Perturbation Quotient
RAP	Relative Average Perturbation

# Capítulo 1

## Introdução

Ao longo da primeira secção deste documento, será feita uma descrição do tema escolhido e dos principais objetivos a serem alcançados. Numa fase seguinte, será indicado um conjunto de abordagens a seguir para os atingir. Será também apresentada a estrutura de todo o documento.

### 1.1 - Descrição do tema

Ao longo deste estudo, pretende estabelecer-se um conjunto de correspondências entre distorções existentes na voz (nomeadamente na distribuição temporal dos seus impulsos glóticos, o jitter) e nos impulsos cardíacos, em situação de repouso. Este tipo de estudo poderá conduzir a um conjunto de conclusões inovadoras sobre a forma como a voz nos diz algo sobre o estado do nosso coração.

Esta avaliação será feita tanto para o mesmo indivíduo, como para um conjunto de indivíduos, possibilitando tirar conclusões individuais e abrangentes relativamente a este conjunto de correspondências.

Importa salientar que um dos parâmetros a avaliar - a frequência fundamental - é mais facilmente detetável se o teste for feito com fonação de vogais sustentadas. No caso em que tal não aconteça, será desencadeado um conjunto de mecanismos de normalização, para que se possa tirar o mesmo tipo de conclusões.

Outros detalhes sobre o tema serão explicados ao longo das próximas secções.

### 1.2 - Motivação

Um dos principais motivos que levou à escolha por este tema em detrimento de outros foi o gosto por processamento de sinal, nomeadamente de sinais de voz. Para além disso, o facto

de, neste caso concreto, este conhecimento se cruzar com outra área de interesse - a Medicina - motivou bastante. Será uma boa oportunidade de cruzar estes dois domínios com um estudo que poderá, numa perspetiva mais filosófica, conduzir a conclusões interessantes relativamente ao modo como os seres humanos interagem, quer estejam a mentir (no contexto de um polígrafo, por exemplo), quer estejam a sentir outro tipo de emoção. Porque a componente emocional é, em larga medida, o que faz de nós humanos.

### **1.3 - Objetivos a atingir**

Como é indicado anteriormente, o principal objetivo é estabelecer um conjunto de ligações entre parâmetros de caracterização ou perturbação de voz e do eletrocardiograma. Para atingir este objetivo será feita, numa fase inicial, uma caracterização detalhada do Estado da Arte. Esta tem por objetivo conferir um conjunto de conhecimento essencial à compreensão do estudo, permitindo também um melhor enquadramento no tema para os leitores da dissertação.

Numa fase seguinte, terá de ser recolhido um conjunto de amostras de sinais de voz e ECG num conjunto de voluntários. Este conjunto de pessoas deverá enquadrar-se, no caso ideal, num lote de perfis relativamente diversificado, de modo a enriquecer as conclusões a serem tiradas posteriormente. A metodologia destes testes será descrita detalhadamente mais à frente. Depois de recolhido o conjunto de amostras, proceder-se-à ao estudo destas distorções recorrendo a software específico, nomeadamente VoiceStudio ou Praat.

Depois desta análise, pretende-se estabelecer correspondências nestas caracterizações de sinal para o mesmo orador (intraorador) e para vários oradores - dispersão interorador. Será tirado um conjunto de conclusões que culminarão, desejavelmente, na publicação de um artigo científico.

### **1.4 - Estrutura**

A estruturação do documento foi feita dividindo-o em cinco secções distintas. O propósito da primeira foi já descrito no parágrafo anterior. No capítulo 2 será feito o Estado da Arte para o tema escolhido, abordando todas as temáticas que ajudem à melhor compreensão do estudo de uma forma muito breve, para além dos avanços relativos ao tema propriamente dito. No capítulo seguinte, serão descritas as tecnologias a utilizar ao longo da dissertação.

O capítulo 4 pretende elucidar relativamente à estruturação e planeamento do projeto. Por último, existe um capítulo para conclusões.



# Capítulo 2

## Estado da Arte

Nesta secção será abordado um conjunto de conceitos que ajudarão à melhor compreensão do tema a abordar. Falar-se-á sobre o exame típico de eletrocardiograma, nomeadamente dos principais eventos, dos riscos associados, do equipamento utilizado e de alguns exemplos de anomalias que poderão ser detetados recorrendo a este teste.

De seguida, será abordado um conjunto de conceitos relacionados com análise de som e voz. Depois de explicada a relação entre a distorção da voz e do eletrocardiograma, serão descritas as tecnologias a utilizar.

### 2.1 - Caracterização do eletrocardiograma

O eletrocardiograma foi inventado por Willen Einthoven, da Universidade de Leiden, na Holanda. Tal invenção valeu-lhe o prémio Nobel da Medicina em 1924. Trata-se de uma análise transtorácica que mede a atividade elétrica no coração, durante um período de tempo, através de um conjunto de elétrodos em contacto com a pele. Esta ferramenta é utilizada para o diagnóstico e tratamento de praticamente todos os tipos de doenças do coração [1.1, Cap.12]. Trata-se de um exame que é fruto de uma série complexa de processos fisiológicos e tecnológicos.

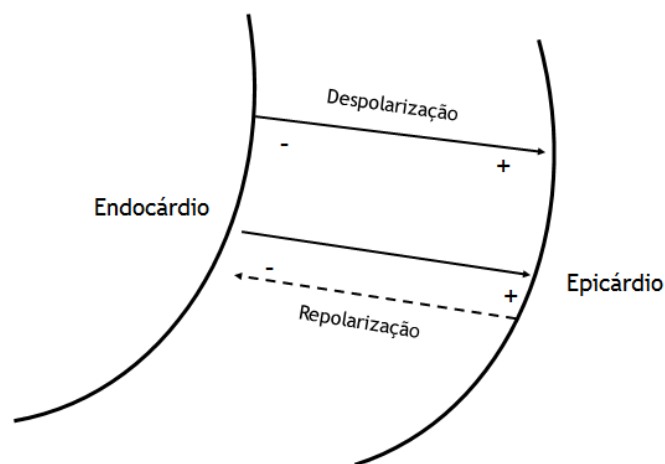
#### 2.1.1 - Enquadramento fisiológico

Inicialmente, formam-se correntes de origem iónica ao longo da membrana celular e das suas células adjacentes. Estas correntes são sincronizadas com a ativação cardíaca e geram um campo elétrico dentro e em torno do coração que varia ao longo de um ciclo cardíaco. Este campo elétrico atravessa outras estruturas do nosso corpo, sendo possível detetar ecos da atividade do coração nestes locais.

## 4 Estado da Arte

As correntes iónicas transmembranares são as principais responsáveis pelas medições gravadas num ECG. Considera-se, de um modo arbitrário, que estas correntes se representam por movimentos de cargas positivas.

Numa situação de repouso, cada célula do músculo cardíaco apresenta uma carga negativa. Esta carga é anulada através de um fluxo de iões com carga positiva, como o Sódio ( $\text{Na}^+$ ) ou o Cálcio ( $\text{Ca}^{2+}$ ) - a este fenómeno chama-se despolarização (ver fig. 2.1). Este evento é o principal responsável pelo desencadear das contrações cardíacas. No caso convencional, esta despolarização é promovida, a cada ciclo cardíaco, pelas células do nó sino-auricular. Esta estrutura é responsável por enviar os impulsos elétricos que percorrem as aurículas e estimulam o nó auriculo-ventricular (que posteriormente transmitirá impulsos aos ventrículos), pelo que tem um papel fundamental no controlo da frequência dos batimentos cardíacos [1.1, Cap. 12].



**Figura 2.1** - Mecanismo de despolarização e repolarização

### 2.1.2 - Aplicabilidade

O exame é muito utilizado em análises de rotina, quando um paciente é admitido nas urgências ou antes de uma intervenção cirúrgica, de modo a detetar anomalias até então desconhecidas. Em todo o caso, existe alguma flexibilidade nos regulamentos quanto à obrigatoriedade do teste nestas situações.

O eletrocardiograma é unanimemente considerado o melhor método de medir e diagnosticar qualquer arritmia cardíaca [1.1, Cap. 12]. Em particular, indica-se para encontrar anomalias no tecido condutor onde os sinais elétricos são transmitidos, ou anomalias nas batidas cardíacas causadas por problemas na despolarização. Nos casos de ocorrência de enfartes do miocárdio é possível, na maior parte das vezes, verificar que zonas do músculo cardíaco se danificaram.

A principal limitação deste tipo de teste é que não consegue medir a capacidade de bombeamento do coração, sendo para isso necessário recorrer a técnicas de ultrassom como o ecocardiograma ou outros testes de medicina nuclear. Estes testes são de enorme importância, uma vez que é possível um ser humano com um sinal de ECG perfeitamente normal entrar em paragem cardíaca, num fenómeno chamado *pulseless electrical activity* [1.1, Cap. 12].

### 2.1.3 - Eventos

Um exame típico de ECG é composto por um conjunto de eventos fundamentais: onda P, complexo QRS, onda T e onda U (esta última visível em cerca de metade dos exames). A linha horizontal do ECG é comumente designada de linha de base. Para além destes elementos, outros são obtidos por análise indireta: intervalos PR, ST ou QT. Alguns destes exemplos encontram-se ilustrados na figura 2.2.

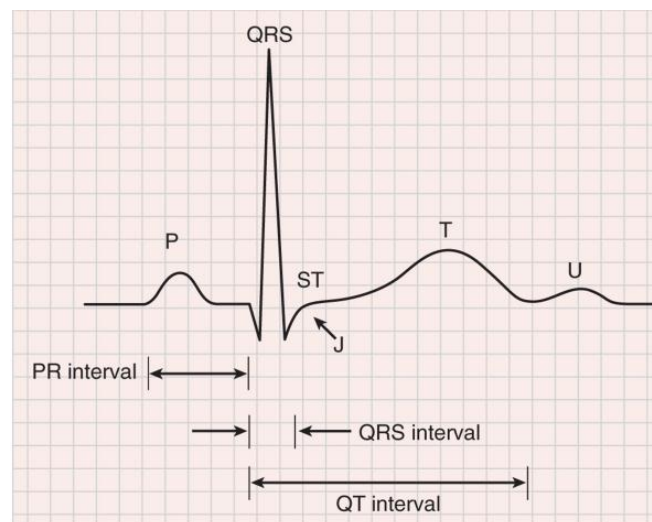


Figura 2.2 - Alguns Eventos de um ECG [1.1]

- Linha de base: ponto de partida para atividade elétrica inerente às despolarizações e repolarizações;
- Onda P: durante a despolarização auricular normal, o vetor campo elétrico está orientado do nódulo sino-auricular para o nódulo aurículo-ventricular, distribuindo-se da aurícula direita para a esquerda, originando a onda P. Apresenta duração típica de 80 ms;
- Complexo QRS: reflete as variações rápidas na despolarização ventricular. Como os ventrículos apresentam maior massa muscular comparativamente com as aurículas, a amplitude da onda será muito maior que na onda P. Apresenta uma duração típica em torno dos 80-120 ms;
- Onda T: representa a repolarização ventricular. Divide-se pelo pico em duas regiões distintas: período refratário absoluto e relativo (ou vulnerável) (duração típica de 160 ms);

## 6 Estado da Arte

- Onda U: repolarização do septo interventricular. Normalmente apresenta amplitude baixa ou quase nula. Caso tal não ocorra, poder-se-à estar perante patologias como hipertireoidismo ou hipercalcemia [1.2].
- Intervalo PR: demarcado pelo início da onda P e início do complexo QRS. Reflete o tempo que o impulso elétrico demora a atravessar desde o nó sinusal até ao nó auriculo-ventricular, antes de entrar nos ventrículos. Tem a duração típica de 120-200 ms;
- Segmento PR: semelhante ao anterior, mas começando no final da onda P. Representa o intervalo entre o fim da despolarização auricular e o início da polarização ventricular. Tem a duração típica de 50 a 120 ms;
- Segmento ST: intervalo entre o final da onda S e início da onda T, representa o intervalo de excitação ventricular. Tem a duração de 80 a 120 ms;
- Intervalo QT: vai desde o início do complexo QRS até ao final da onda T, representando o tempo de sístole elétrica. A sua medição pode ser importante para prevenir taquiarritmias ou morte súbita.

São detetáveis outros eventos na medição do ECG. No entanto, estes são considerados os mais importantes, sendo capazes de cobrir uma quantidade de medidas suficiente para o estudo em questão.

### 2.1.4 - Principais riscos

Como todos os outros exames médicos, o ECG apresenta uma margem de erro. A eficácia deste exame como ferramenta de diagnóstico está dependente de fatores como as indicações existentes para o procedimento, o contexto clínico no qual é feito, o uso correto dos dispositivos e a capacidade do examinador. O ruído no sinal devido às forças que atuam nos elétrodos, especialmente às altas frequências, pode ser outro fator de erro.

A análise do ECG feita sem conhecimento do historial clínico do paciente pode ser limitativa. Torna-se relevante, por exemplo, saber que medicação está a ser consumida, uma vez que esta pode adulterar os resultados. A existência de registos de ECG anteriores pode também ter utilidade.

Outro problema comum que pode levar a erros de diagnóstico está relacionado com erros técnicos (por exemplo, troca de elétrodos). Situações destas podem conduzir a posteriores testes e tratamentos que podem colocar o paciente em perigo ou a desperdício de recursos. Conhecer os padrões de leitura deste tipo de erro ajuda no seu despiste.

Erros na interpretação de ECGs têm aumentado nos últimos anos. Estes erros culminam em falhas na deteção e triagem de pacientes com doenças que colocam em jogo a vida dos pacientes. Estudos concluem que erros de grande proporção podem ser identificados em 4 a 32% das interpretações [1.1, Cap. 12]. Além deste tipo de erros, devem-se ter em conta os erros computacionais. Apesar de os computadores serem uma ajuda importante em termos de

processamento e algoritmia, os instrumentos de medida ainda apresentam limitações que impedem deteção de problemas mais complexos. Novas técnicas baseadas em inteligência artificial, aliadas a avanços no hardware, constituem um avanço que tentará colmatar esta falha.

### 2.1.5 - Dispositivos utilizados

Para realizar o exame, é necessário um conjunto de elétrodos para captar as diferenças de potencial. É também necessário um sistema de visualização (ver figura 2.3). A qualidade desta visualização está dependente do tipo de sistemas eletrónicos usado para amplificação, filtragem e digitalização da informação obtida. Depois deste tratamento de informação, o registo é normalmente imprimido num papel quadriculado com dimensões próprias, num gráfico de voltagem por tempo. Antes da medição, é prática comum a obtenção de um sinal de calibragem.

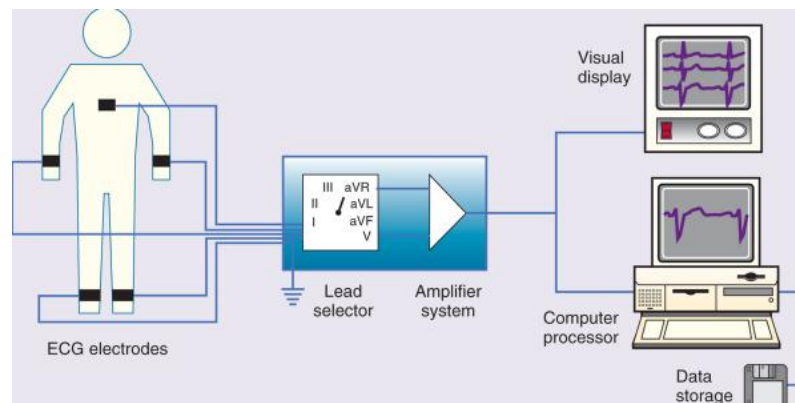


Figura 2.3 - Principais dispositivos de um ECG [1.1]

A portabilidade e consequente aceitação dos dispositivos de medição de ECG, especialmente em pacientes de idade mais avançada, deve ser um fator a ter em conta no seu projeto [1.3].

### 2.1.6 - Exame típico

O exame mais comum é o ECG de 12 derivações. Cada derivação corresponde a uma diferença de potencial entre dois elétrodos colocados em zonas diferentes do corpo. Antes de iniciar o procedimento, o examinador deverá explicar as etapas do processo. A sala deverá ter uma temperatura agradável e o paciente deverá estar descansado antes do procedimento. Deve-se também conhecer a sua medicação [1.1, Cap. 12].

Depois de colocar o paciente em decúbito dorsal (barriga para cima), deve determinar-se a posição das derivações. Normalmente, quatro irão para os membros e as seis restantes para posições do peito. De seguida, é colocado o gel de condução nos respetivos locais, ligando-se

os elétrodos ao aparelho de medição. Pode ser necessária uma tricotomia (corte de pelos) no caso masculino. Os sinais são depois visualizados e registados no papel quadriculado por quem os interpretará convenientemente [1.1, Cap. 12].

### 2.1.7 - Exemplos de anomalias

Por análise dos eventos anteriormente descritos torna-se possível, para um profissional treinado, fazer um diagnóstico diretamente a partir da leitura do ECG. Por exemplo, um intervalo QT prolongado/encurtado pode indicar hipercalcemia ou uma anomalia genética. Uma onda T invertida ou achatada, isquemia coronária ou hipertrofia no ventrículo esquerdo. Uma onda T hiperaguda pode ser indicativa de um futuro enfarte no miocárdio. Uma extrassístole ventricular é diagnosticada graças à presença de um complexo QRS prematuro e de morfologia larga ou deformada com duração superior a 120 ms, seguido de uma pausa compensadora.

Como estes, existem muitos outros exemplos de variações nos eventos que apontam para um diagnóstico, bem como mecanismos automáticos para classificação de vários tipos de arritmias [1.4]. No entanto, dado que este estudo não engloba a vertente de diagnóstico, não se aprofundará mais esta abordagem.

## 2.2 - Caracterização de parâmetros do som

O som é uma onda mecânica que representa uma oscilação de pressão transmitida por um meio físico (sólido, líquido ou gasoso), composto por uma ou mais frequências. Estas frequências podem ou não estimular os mecanismos de audição, isto é, serem ou não serem audíveis. Neste sub-capítulo apresenta-se um conjunto de parâmetros caracterizadores do som que auxiliam na sua análise acústica.

### 2.2.1 - Intensidade sonora

É a percepção da amplitude da onda sonora. A percepção desta grandeza pelo ouvido humano apresenta uma característica logarítmica. Isto implica que variações apenas são percecionadas no caso das amplitudes variarem exponencialmente. A medida de pressão sonora em relação à percepção auditiva é feita recorrendo a uma unidade logarítmica, o decibel (dB) [1.6, Secção 2.2].

### 2.2.2 - Frequência fundamental (F0) e formantes (F1, F2... Fn)

A frequência fundamental corresponde ao primeiro harmónico de uma onda sonora, sendo a frequência mais influente na percepção de um dado som. No caso da voz, existe variação

deste valor consoante a idade e o sexo. Num homem adulto está entre os 85 e os 180 Hz, e na mulher entre os 165 e os 255 Hz. É um dos principais elementos caracterizadores da voz.

Os formantes são os picos seguintes no espectro sonoro. Trata-se de uma informação distintiva importante na voz falada e cantada. O primeiro formante apresenta a menor frequência, e assim sucessivamente. Os dois primeiros formantes (F1 e F2) costumam ser suficientes para identificar uma vogal sustentada [1.6, Secção 2.2].

### 2.2.3 - Tonalidade, timbre e intensidade subjetiva

Na acústica, a tonalidade é reconhecida pela frequência de vibração do som, ajudando à distinção entre graves e agudos. Um tom simples é composto por apenas uma frequência, e um complexo por várias, sendo que a mais baixa indica o tom fundamental. As frequências dos tons seguintes são múltiplos da sua frequência fundamental. Esta combinação agradável de tons compõe um tom musical [2.1].

O timbre é a característica de um tom que o distingue dos restantes num contexto de produção sonora. Dois instrumentos que produzam a mesma nota musical ao mesmo volume apresentam características diferentes no domínio das frequências, logo diferentes timbres.

A intensidade subjetiva (ou *loudness*) relaciona alguns dos parâmetros anteriormente descritos com a perceção qualitativa do ouvido humano. Pode variar de pessoa para pessoa, pelo que não é possível atribuir-lhe uma métrica fixa. Este parâmetro é frequentemente confundido com a intensidade sonora.

### 2.2.4 - Representação e análise de Fourier

Para fazer análise dos sons, é necessário convertê-los primeiro para o domínio das frequências. Neste domínio, é comum utilizar representações gráficas do módulo e da fase do sinal. Para uso computacional é frequentemente utilizada a FFT (Fast Fourier Transform), que é uma discretização da Transformada de Fourier (DFT - Discrete Fourier Transform).

### 2.2.5 - Espectro e espectrograma

Os sinais aos quais é aplicada a transformada são representados em espectros. Todos os sinais com variação temporal são representáveis desta forma. Esta representação torna os harmónicos mais visíveis, ajudando a uma análise mais eficaz do sinal. Decompor um sinal complexo em partes mais simples constitui uma prática comum neste tipo de análise, especialmente em sinais periódicos.

O espectrograma representa a variação temporal de um espectro de frequências, através de um gráfico com um eixo horizontal que representa o tempo, um vertical que representa a frequência e variações na cor que indicam a amplitude de uma dada frequência.

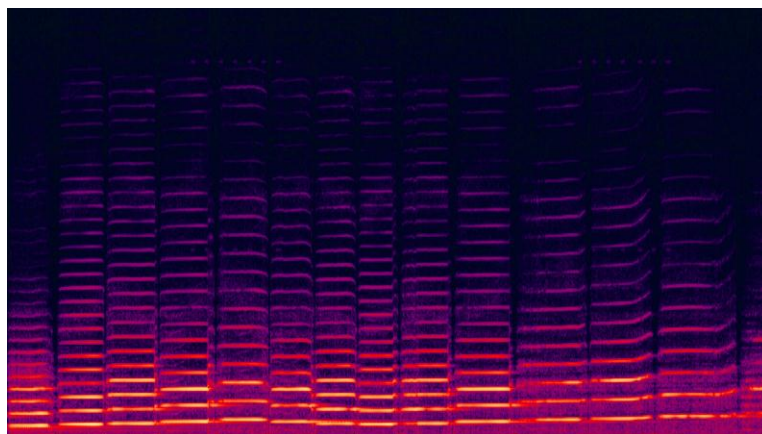


Figura 2.4 - Espectrograma de um violino [2.2]

## 2.3 - Produção e Percepção da Fala

Após caracterização dos parâmetros fundamentais do som, importa perceber quais os mecanismos desencadeados pelo ser humano para os produzir/ouvir.

### 2.3.1 - Percepção auditiva

Compreender o modo como são percebidos os sons implica algum conhecimento da anatomia do ouvido. De um modo simplificado, este é composto pelo ouvido externo, onde as ondas sonoras são reflectivas e atenuadas, fornecendo a informação necessária ao cérebro sobre a fonte do som. No canal auditivo, os sons entre 3 e 12 kHz são amplificados. Depois de atingida a membrana timpânica, o som atinge o ouvido médio. Aqui, atravessa um conjunto de ossículos (martelo, bigorna e estribo) que aumentam a pressão das vibrações para preparar a etapa seguinte.

O ouvido interno é fundamentalmente constituído pela cóclea, que contém os terminais nervosos responsáveis pela audição. Trata-se de um tubo ósseo espiral composto por três compartimentos membranares e um compartimento central que contém as células que ajudam a conferir a sensação de audição através dos movimentos dos fluídos no seu interior. [1.6, Secção 3.3]



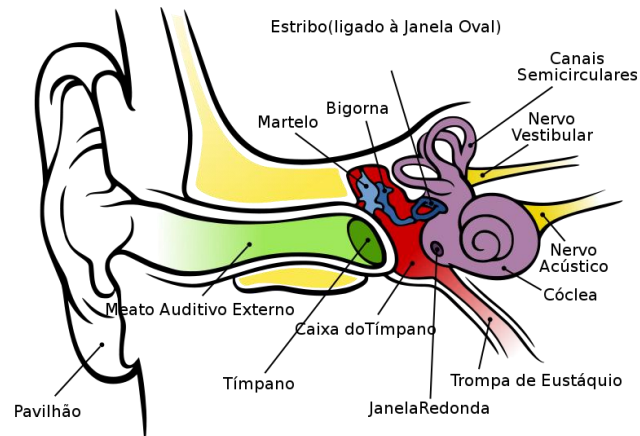


Figura 2.5 - Anatomia do ouvido humano [1.1]

### 2.3.2 - Produção de fala

Define-se produção de fala como o processo a partir do qual palavras faladas são escolhidas e articuladas recorrendo ao aparelho vocal humano. Normalmente, a fala é criada através da pressão promovida pelos pulmões, gerando som por fonação na glote. Consoante a palavra a articular, a vibração das cordas vocais será diferente. Consideram-se três fases de processamento na fonação: conceptualização (no qual se define a intenção de criar o fonema), criação do *lemma* (palavra sem som associado) e articulação (que envolve todas as mudanças físicas inerentes à fala). [1.5]

### 2.3.3 - Modos e lugares de articulação

O modo como a fala é produzida depende do modo como o ar chega ao exterior. No caso de se tratar de um fluxo periódico de ar por vibração das cordas vocais, estamos perante fala vozeada. Em contraste, a fala não vozeada abdica deste mecanismo, tornando difícil a distinção entre alguns fonemas (por exemplo, /f/ e /v/).

No caso das vogais, o local de articulação caracteriza-se pela posição da língua, pela abertura maxilar ou pelo arredondamento da boca. No caso das consoantes, a única informação disponível é o local da constrição, normalmente supra-glótica. [1.6, Secção 3.2]

### 2.3.4 - Mecanismos de fonação

O tipo de som que produzimos pode ser classificado de acordo a disposição das pregas vocais. Se estiverem imóveis, considera-se a situação de fonação não vozeada. Se a glote for comprimida, existe um ponto intermédio onde a vibração das cordas vocais é máxima - este ponto define a voz modal presente nas vogais [1.6, Secção 3.2].

Existe também a voz murmurada, quando existem porções sem fonação (por exemplo, na palavra “japonês”). Outros exemplos poderiam ser considerados, como a voz laringelarizada, nasalada ou de falsete.

### 2.3.5 - Modelo Fonte-Filtro

Para o estudo dos mecanismos anteriormente descritos é muito utilizado o modelo fonte-filtro, em que a fonte são os impulsos/ruídos glotais e o filtro as influências das ressonâncias do trato vocal (oral e nasal). Por ser uma aproximação simplificada e precisa, é utilizada nas mais variadas aplicações, nomeadamente em compressão áudio.

Como os impulsos glotais têm um comportamento periódico, o seu espectro será um conjunto de harmónicos da frequência fundamental cuja magnitude decresce a 12 dB por oitava. A esta fonte é aplicado um filtro que, no caso mais simples, será composto unicamente por pólos cujos coeficientes são obtidos através de mecanismos de predição linear [1.6].

## 2.4 - Medição de parâmetros de voz

Depois de compreender os mecanismos que geram e captam a voz, importa introduzir um conjunto de medidas que a avaliam. Estas medidas poderão ter em conta vários critérios.

### 2.4.1 - Critérios de avaliação

No caso de se tratar de uma avaliação fisiológica, exames como a laringoscopia ou a estroboscopia são os mais indicados. Numa avaliação perceptiva, existe um conjunto de protocolos que ajudam a classificar uma voz. No entanto, a sua elevada subjetividade conduz a análises algo inconsistentes. [1.6].

O critério que apresenta maior número de vantagens é a avaliação acústica. Nele existe um conjunto de parâmetros objetivos que permitem uma comparação documentada e quantificada. Nesta avaliação promove-se o uso de fonação sustentada, captando um mínimo de cem impulsos glotais para avaliação. Avalia-se também a frequência fundamental (número de impulsos glóticos por segundo) em termos de média e desvio padrão.

É frequente o uso de avaliação acústica e perceptiva simultaneamente, de modo a encontrar modelos que as correlacionem. Este tipo de abordagem é ainda alvo de intensa investigação

### 2.4.2 - Jitter

O jitter é a variabilidade involuntária do período fundamental de um sinal de voz, ciclo a ciclo. Existem vários métodos para o seu cálculo, entre os quais o RAP (Relative Average Perturbation) ou o PPQ (Period Perturbation Quotient). Está relacionado com a aspereza da voz. Para uma gama normal em vogal sustentada (> 100 ciclos), o RAP não deverá ser superior a 1.5%.

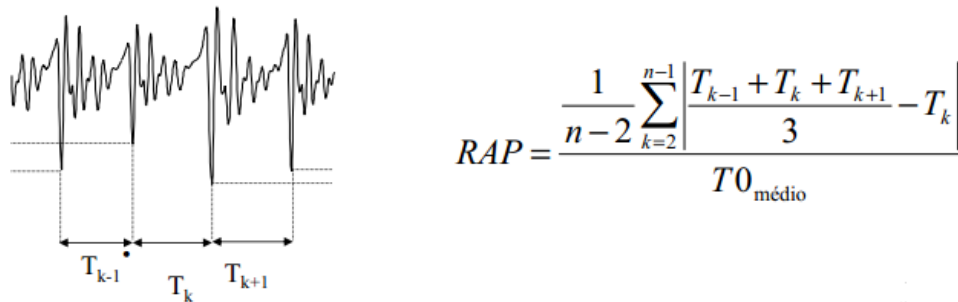


Figura 2.6 - Representação de impulsos glotais (à esquerda) e fórmula de cálculo do RAP (à direita)

### 2.4.3 - Shimmer

O shimmer é a variabilidade involuntária da amplitude dos ciclos glotais, estando relacionado com a aspereza da voz. Uma das fórmulas de cálculo é o APQ (Amplitude Perturbation Quotient). Para as condições referidas anteriormente este valor, para onze ciclos na sua vizinhança, não deverá ultrapassar os 3%.

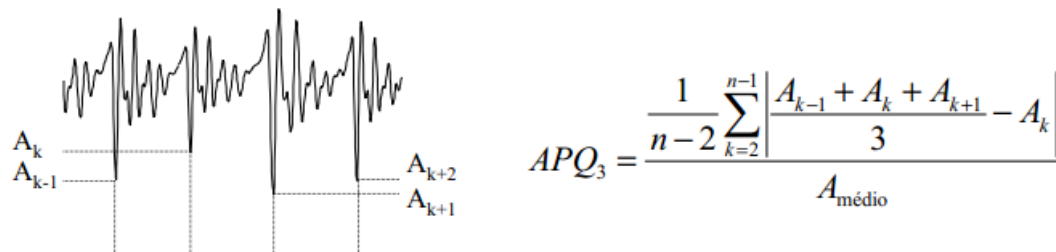


Figura 2.7 - Representação de impulsos glotais (à esquerda) e fórmula de cálculo do APQ (à direita)

### 2.4.4 - Harmonic-to-noise ratio (HNR)

Relação entre a energias das componentes harmónicas e das componentes de ruído do sinal de voz limpo. Exprime-se em dB para se aproximar melhor da apreciação perceptiva, por humanos, da qualidade da voz. Para as condições referidas anteriormente, deve ser superior a 12 dB. Está maioritariamente relacionado com a soproiedade e astenia.

$$HNR = 10 \log_{10} \frac{\text{energia}_{\text{harmónica}}}{\text{energia}_{\text{ruído}}} = 10 \log_{10} \frac{\int_f |H(f)|^2}{\int_f |N(f)|^2}$$

Figura 2.8 - Fórmula para o cálculo do HNR

#### 2.4.5 - Tempo máximo de fonação

Este parâmetro ajuda a medir a eficiência glótica. Mede o tempo durante o qual se dá a emissão de uma vogal, com intensidade constante e confortável. Esta medição é feita por duas ou três vezes, obtendo-se o melhor resultado. Caso este valor seja inferior a 10 segundos, pode-se estar perante um quadro patológico.

#### 2.4.6 - Estimação de impulsos glotais (EGG)

Este mecanismo é de enorme importância tendo em conta a modelização destes impulsos para síntese artificial da voz. A estimação é feita recorrendo a uma electroglotografia. Neste teste, mede-se a resistência elétrica das estruturas laríngeas, colocando simetricamente dois elétrodos nas cordas vocais. As formas de onda obtidas indicam as fases de abertura e fecho das pregas vocais.

Este método é imune a ruído, uma vez que não é influenciado pelo trato vocal - a medição é feita diretamente na fonte. A deteção de desvios na posição e no tempo de abertura/fecho nas pregas vocais pode relacionar-se com certas patologias.

Entre as limitações do teste está a dificuldade de captação devido a uma deficiente colocação dos elétrodos, muito por culpa das variações na estrutura da laringe dos examinados.

## 2.5 - Avanços na avaliação de distorções na voz e no ECG

Neste capítulo será feito um levantamento de alguns avanços feitos na relação entre parâmetros de distorção de voz e ECG.

### 2.5.1 - Modelos Existentes

Numa investigação feita por Dmitriy E. Skopin e Sergey U. Baglikov, é provado que a modulação dos impulsos cardíacos é possível para a voz humana e que essa informação pode ser extraída no domínio das frequências [1.8]. Esta extração é feita recorrendo ao espectro bidimensional proveniente da fonação de uma vogal, considerando um modelo fonte-filtro explicado na secção 2.3.5.

Tendo em conta que a laringe contém músculos cobertos por vasos sanguíneos, torna-se igualmente possível relacionar as batidas cardíacas com as variações dinâmicas de parâmetros da voz como o volume ou tempo de fonação. O sistema de captura consiste num microfone e um sistema portátil de captura de ECG. É previsto que o som captado seja posteriormente transmitido.

Existem também avanços no que diz respeito à identificação de emoções a partir da voz e do ECG. A duração dos ciclos cardíacos é feita recorrendo a parâmetros da fala, e os formantes F2 e F3 identificam vários estados emocionais. No caso deste estudo, descobriram-se os efeitos que a raiva e o medo têm no coração. Para obter a relação pretendida, os sinais de voz e ECG foram capturados simultaneamente [1.9].

Num outro contexto, foi levado a cabo um estudo que indicava que a fonação em pacientes que sofrem de apneia do sono é diferente daqueles que não apresentam esta patologia. Para chegar a esta conclusão foram recolhidas amostras de som de um conjunto de voluntários com um perfil específico [1.10]. Em oposição à análise convencional utilizada neste caso, é também possível seguir uma análise não linear dos sinais de voz, associando-os a vozes saudáveis ou patológicas [1.11].

## **2.6 - Conclusões**

De um modo geral, o estudo de distorções na voz e no ECG e a sua comparação em registos recolhidos na mesma janela temporal é uma temática que, tanto quanto a pesquisa efetuada conseguiu apurar, está relativamente inexplorado. Não quer isso dizer que a informação recolhida seja inválida, uma vez que confronta um conjunto de perspetivas diferentes que conferirão um conjunto de conhecimentos e abordagens que certamente se revelarão úteis.

# Capítulo 3

## Ferramentas a utilizar

Depois de identificada a principal informação de base para o desenrolar do projeto, importa escolher quais as ferramentas a ser utilizadas para levar a cabo a prova do conceito. Naturalmente que se trata de uma lista que poderá sofrer pequenas alterações. No entanto, as principais ferramentas estão já identificadas.

### 3.1 - Algoritmia de Processamento de Sinal

Para efetuar as medições e comparações mencionadas ao longo deste Estado da Arte, é necessário recorrer a um conjunto de algoritmos que lidam com análise de sinal. Neste caso, os sinais a analisar são biológicos (ECG) e, como tal, contínuos no tempo.

Será prática corrente converter os sinais para o domínio das frequências e fazer uma análise espectral, pelos motivos anteriormente referidos. O software a ser utilizado para o efeito será o Matlab, que está especialmente vocacionado para cálculo numérico e possui várias ferramentas para análise de sinais em geral. Além disso, existe uma um conjunto de ferramentas especialmente direcionadas à análise de sinais de voz - a *Speech Processing Toolbox*.

Para auxiliar ao reconhecimento e classificação de doenças cardíacas, existem simuladores de ECG que tiram partido da elevada capacidade de processamento do Matlab. Este tipo de abordagem é muito utilizado em contextos de aprendizagem. [1.12].

### 3.2 - Praat

Palavra holandesa para “falar”, o Praat é um software livre para análise de sinais de fala. Foi desenvolvido por Paul Boersma e David Weenick da Universidade de Amsterdão. A sua

versão mais recente é a 5.1, datada do final de janeiro de 2009. Apesar de ser um programa poderoso, o seu uso para prática profissional pode não ser o mais apropriado.

### 3.3 - SEEGNAL Voice Studio

Também direcionado para processamento de sinais de voz, oferece um conjunto de ferramentas intuitivas e expeditas numa interface amigável. Neste software é possível a gravação e escrita instantâneas utilizando um microfone em formatos comprimido e não comprimido.

É também possível a obtenção dos principais parâmetros objetivos de qualidade da voz, nomeadamente o F0, jitter, shimmer, HNR e energia - todos eles descritos anteriormente. Para além disto, destaca-se o facto de estas funções de análise conseguirem correr durante a captação/reprodução.

### 3.4 - Biopac MP36RWSW

Este será o dispositivo utilizado para a captação dos ECG. É também capaz de efetuar outros tipos de captação, como eletromiografias ou eletroencefalografias. Possui quatro canais para aquisição de dados, e é acompanhado de um software interativo e intuitivo para visualizar as medições em tempo real. Tem diversas opções de visualização, bem como capacidade de exportar informação para ser usada noutros contextos.

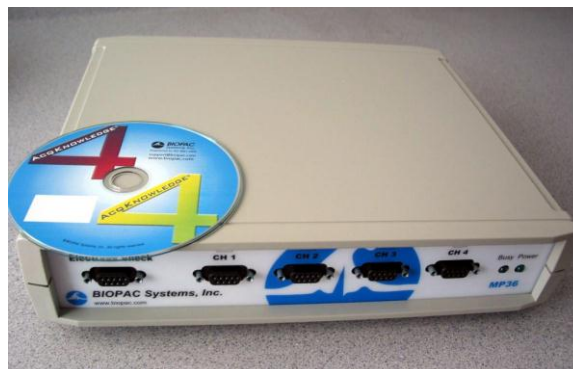


Figura 3.1 - Aparelho Biopac [2.3]

# Capítulo 4

## Estruturação do Projeto

Nesta secção será dado um conjunto de informações sobre o projeto propriamente dito. Inicialmente, indica-se o conjunto de resultados esperados no seu final, seguido dos principais desafios. Por último, apresenta-se o plano de trabalho a seguir até ao final da dissertação.

### 4.1 - Resultados esperados

Depois de uma investigação cuidada para familiarização com o tema e com todos os conceitos a ele associados, espera-se que seja feita uma recolha de amostras ECG em voluntários. Depois de se obterem estas amostras, passar-se-á a uma fase de análise dos parâmetros de distorção de voz, usando os ambientes anteriormente mencionados.

Espera-se que, após esta análise, seja possível caracterizar a natureza destas variações para o mesmo orador e para conjuntos de oradores. Desde modo, deverá ser possível associar características específicas da voz (eventualmente audíveis) a perturbações no ECG. Este tipo de conclusão apenas pode ser tirado se os intervalos de tempo de análise coincidirem.

### 4.2 - Principais desafios

O caso mais simples para o qual se propõe estabelecer esta comparação é o regime sustentado no caso da voz (produção de duas vogais, /a/ e /i/ por exemplo) e de repouso no caso do ECG. No entanto, um dos principais desafios consiste em tirar conclusões considerando variações lentas de sinal em que se terá de fazer uma modelização da sua evolução média.



Outro dos desafios é conseguir um conjunto de amostras significativo e diversificado, de modo a conseguir tirar conclusões válidas para um maior número de perfis.

### 4.3 - Plano de Trabalho

De seguida apresenta-se o plano de trabalho previsto, à data de entrega deste relatório.

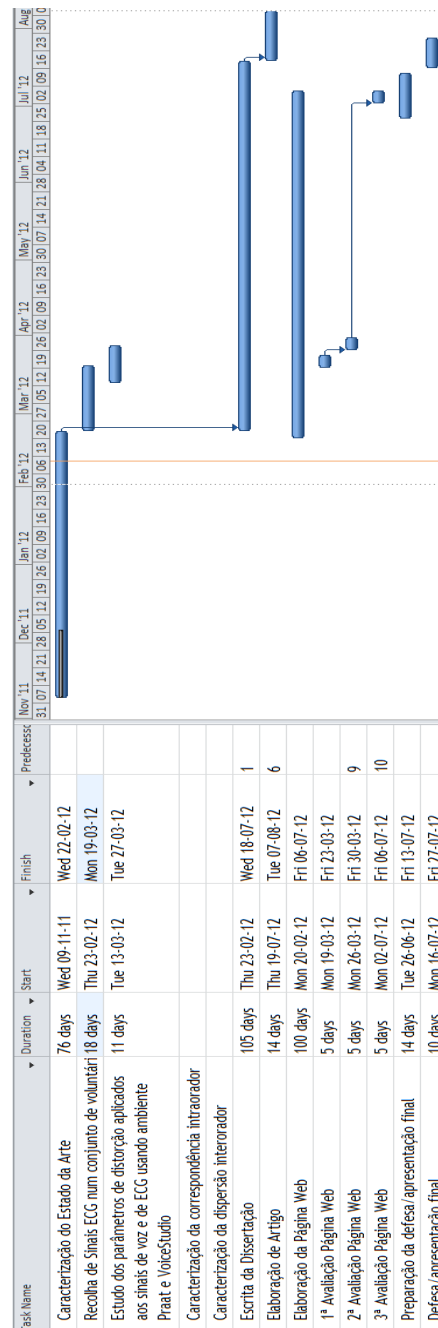


Figura 4.1 - Diagrama de Gantt do Projeto

# Capítulo 5

## Conclusão

Este primeiro documento será verdadeiramente útil na medida em que, para além de conter a informação base para a compreensão do tema desta dissertação, ajudou a compreender qual o estado de desenvolvimento dos modelos em estudo. Todo este exercício conduziu a uma reflexão mais profunda sobre a pertinência do tema, confirmando-a e motivando para o trabalho futuro.

Uma vez que a temática do processamento de voz já tinha sido abordada em Unidades Curriculares anteriores, a pesquisa que consumiu mais tempo foi aquela que teve a ver com o eletrocardiograma. Para o seu sucesso contribuiu decisivamente o livro de cardiologia utilizado, “Braunwald's Heart Disease” [1.1].

A definição do planeamento levantou um conjunto de pequenas dúvidas em algumas fases do projeto, nomeadamente na fase de recolha dos sinais de voz e de ECG. No futuro, importará definir um conjunto de pormenores logísticos relacionados com o local de recolha de sinais, o número e perfil dos voluntários, o tipo de microfones a utilizar, entre outros.

Em todo o caso, esta primeira fase da dissertação correu conforme o esperado, estando lançadas as bases para o trabalho futuro e a entrada na Dissertação em “regime permanente”.

## Referências bibliográficas

- [1.1] BRAUNWALD, Eugene et al., “Braunwald’s Heart Disease”, Eight edition, Elsevier Inc., 2008.
- [1.2] HOUGHTON, Andrew et al., “Making Sense of the ECG - Cases for self assessment”, Hodder Arnold, 2009.
- [1.3] EHMEN, Hilko et al., “Comparison of four different mobile devices for measuring heart rate and ECG with respect to aspects of usability and acceptance by older people”, Journal of Applied Ergonomics, 2012.
- [1.4] CASTILLO, Oscar et al., “Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbors and neural networks combined with a fuzzy system”, Expert Systems with Applications - An International Journal, 2011
- [1.5] JESCHENIAK, Jorg et al., “Word Frequency Effects in Speech Production: Retrieval of Syntactic Information and of Phonological Form”, Journal of Experimental Psychology: Learning, Memory and Cognition, Vol. 20, No. 4, pp. 824-843, 1994
- [1.6] PEREIRA, Fernando et al., “Comunicações Audiovisuais - Tecnologias, Normas e Aplicações”, 1ª Edição, IST Press, 2009
- [1.7] SKOPIN, Dmitriy et al., “Heartbeat Feature Extraction from Vowel Speech Signal Using 2D Spectrum Representation”
- [1.8] DESHPANDE, Nivedita et al., “Assessment of systolic and diastolic cycle duration from speech analysis in the state of anger and fear”, 2012
- [1.9] MORERA, José et al., “Acoustic Analysis of Vowel Emission in Obstructive Sleep Apnea”, October 1993
- [1.10] HENRÍQUEZ, Patricia et al., “Characterization of Healthy and Pathological Voice Through Measures Based on Nonlinear Dynamics”, IEEE Transactions on Audio, Speech and Language Processing, Vol. 17, No.6, agosto 2009
- [1.11] GUNEY, Esra et al., “WebECG: A novel ECG simulator based on MATLAB Web Figure”, Advances in Engineering Software, 2011

## Referências sitigráficas

- [2.1] “tone (sound) - Britannica online encyclopedia”. Disponível em <http://www.britannica.com/EBchecked/topic/599085/tone>. Acesso em fevereiro/2012
- [2.2] “Spectrogram - Wikipedia, the free encyclopedia”. Disponível em <http://en.wikipedia.org/wiki/Spectrogram>. Acesso em fevereiro/2012
- [2.3] “MP36RW | data acquisition and analysis system with AcqKnowledge for MP36R | BIOPAC - Research”. Disponível em <http://www.biopac.com/data-acquisition-analysis-system-mp36r-system-windows>. Acedido em fevereiro/2012