

A medida HNR: sua relevância na análise acústica da voz e sua estimação precisa

José Lopes(1,2), Susana Freitas(3), Ricardo Sousa(1), Joaquim Matos(2), Filipe Abreu(2), Aníbal Ferreira(1,2)

(1)Fac. Engenharia, Universidade Porto, (3)Universidade Fernando Pessoa, (2)SEEGNAL Research, Lda.

Resumo: O Harmonics-to-Noise Ratio (HNR) é um dos parâmetros objectivos mais relevantes de avaliação acústica da voz podendo ser extraído de forma expedita e não-invasiva. Nesta comunicação pretende-se destacar a relevância do HNR, apresentar alguns métodos usados para a sua estimação e demonstrar uma nova abordagem para a estimação e separação precisa das componentes harmónica e de ruído da voz vozeada e sustentada. Demonstra-se o desempenho da nova abordagem usando voz sintetizada com valores pré-definidos de HNR e utilizando, para referência, estimativas obtidas com técnicas alternativas. Aborda-se também a correlação entre o parâmetro HNR e a avaliação perceptiva de vozes disfónicas.

1. Introdução

A avaliação da voz é uma das componentes principais do diagnóstico vocal e precede a intervenção terapêutica. Normalmente é realizada de acordo com um protocolo contendo duas componentes: a avaliação de acordo com parâmetros perceptivos, também designada de avaliação perceptiva, e a análise de acordo com parâmetros objectivos, também designada de avaliação acústica.

No primeiro caso, o especialista, tipicamente Terapeuta da Fala ou Médico Otorrinolaringologista, aprecia as características sonoras da voz do falante, por exemplo em resultado da fonação sustentada de uma vogal, em relação a referências perceptivas, adquiridas pelo especialista durante a sua formação ou exercício profissional, de vozes categorizadas como normais. Há inclusivamente alguns procedimentos de avaliação padronizados que permitem quantificar a severidade das perturbações percepcionadas. É disso exemplo a escala GRBAS (G – avaliação global da disfonia (*grade*); R – rouquidão (*roughness*); B – soproiedade (*breathiness*); A – astenia (*asteny*); S – tensão (*strain*) ou RASAT (Rouquidão, Aspreza, Soproiedade, Astenia, Tensão) (Guimarães 2007).

As siglas indicadas correspondem, do ponto de vista anatomofisiológico (Pinho 2002):

- **Rouquidão:** irregularidade de vibração das pregas vocais. Assim, a voz é percepcionada com ruídos adventícios produzidos a baixa frequência. Este parâmetro verifica-se em casos de: fenda glótica, presença isolada de uma

alteração orgânica ou fenda de qualquer dimensão com alterações da mucosa das pregas vocais (exemplo: nódulos, pólipos ou edemas).

- **Aspereza:** rigidez da mucosa que também causa alguma irregularidade vibratória, especialmente se associada a fenda ou outras alterações, por exemplo edema das pregas vocais (Edema de Reinke). A voz é seca, sem projecção, com ruídos nas altas frequências pela diminuição da onda mucosa. Exemplo: *sulcos glottidis*, quistos e lesões neoplásicas (cancro).
- **Soprosidade:** presença de ruído de fundo, audível, que corresponde fisiologicamente à fenda glótica (abertura entre as pregas vocais).
- **Astenia:** relacionada com o mecanismo de hipofunção das pregas vocais e reduzida energia de emissão do som. Exemplo: *miastenia gravis* ou outras perturbações neurológicas do controle vocal.
- **Tensão:** associada a esforço vocal por aumento da adução glótica (hiperfunção), geralmente inerente ao aumento da actividade da musculatura extrínseca da laringe, com elevação desta. Exemplo: disfonia espasmódica e síndromes de abuso vocal com conseqüente alteração da mucosa (i.e. nódulos ou pólipos).

Os parâmetros avaliados são classificados numa escala de 4 pontos: 0= normal ou ausência de alterações; 1= ligeiro ou discretas modificações; 2= moderado ou alterações evidentes; 3= severo/grave ou com variações extremas. São também contemplados valores intermédios. Esta é uma escala de triagem vocal que se debruça sobre a fonte glótica durante a produção de vogais sustentadas (/a/ ou /ε/) ou fala encadeada (Pinho 2002).

Contudo, por natureza, a avaliação perceptiva é intrinsecamente subjectiva e, como tal, é caracterizada por uma incerteza que depende não só da referência de vozes normais assumida por cada especialista, como também do grau de severidade subjectivamente avaliado. Em conseqüência, é provável alguma dispersão na avaliação subjectiva da mesma voz quando efectuada por diferentes especialistas¹.

No segundo caso, o sinal acústico decorrente do vozeamento de uma vogal sustentada (tipicamente o /a/) é captado por um microfone, digitalizado e, posteriormente, analisado através de um procedimento computacional para medição de factores de perturbação objectivos, associados à forma de onda do vozeamento. A Figura 1 ilustra este processo. O procedimento computacional recorre a técnicas de Processamento Digital de Sinal (PDS) que permitem a medição de parâmetros objectivos como a frequência fundamental da voz (ou *pitch*²), parâmetros de perturbação como o *jitter* ou *shimmer*, ou parâmetros de qualidade como a relação harmónicos-ruído (HNR).

¹ E até pelo mesmo especialista em momentos diferentes.

² Em rigor, o *pitch* denota o correspondente psicofísico (i.e., perceptivo) da frequência fundamental (F_0) e é condicionado por outros factores objectivos do sinal de voz como seja a sua intensidade. Contudo, para simplificar a discussão, consideramos neste artigo que *pitch* e F_0 são sinónimos.

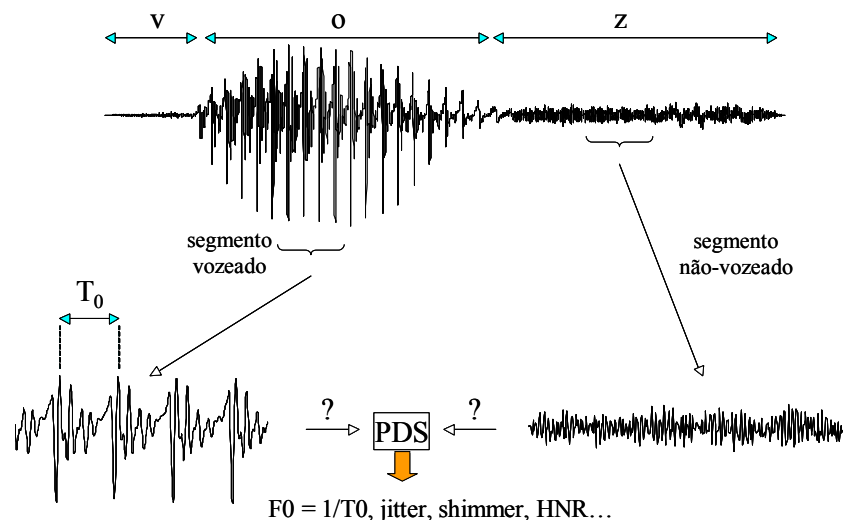


Figura 1: Ilustração do sinal de voz captado por um microfone e correspondente à palavra *voz*. Destaca-se a região vozeada do sinal e a região não-vozeada. Usando técnicas de Processamento Digital de Sinal é possível a medição objectiva e precisa de alguns parâmetros de perturbação extraídos directamente do sinal acústico.

Apesar de objectivos, os parâmetros acústicos não substituem a avaliação perceptiva por duas razões fundamentais. Por um lado, as dimensões de apreciação de qualidade de uma voz, quando avaliada perceptivamente, são em maior número (e portanto mais ricas) do que o número de parâmetros acústicos relevantes que reúnem maior consenso e aceitação na comunidade científica. Por outro lado, a correlação entre os parâmetros acústicos e os perceptivos é ainda matéria de investigação e debate na comunidade científica, o que denota a dificuldade clássica que existe em exprimir a acuidade auditiva humana através de modelos matemáticos. Neste contexto, os parâmetros acústicos representam uma componente importante da avaliação da voz, complementando o diagnóstico e reforçando a avaliação perceptiva.

Este artigo incide sobre um dos parâmetros acústicos mais importantes e aceites que é a relação HNR. Na secção 2 define-se este parâmetro, descrevem-se algumas técnicas para a sua medição e apresenta-se um novo método de cálculo directo e preciso. Na secção 3 caracteriza-se o desempenho deste novo método usando voz sintética e voz natural, em relação a técnicas alternativas. Aborda-se também a correlação entre o parâmetro HNR e a avaliação perceptiva. Na secção 4 resumem-se as principais conclusões.

2. A medida HNR

2.1. Definição

A medida HNR é uma avaliação objectiva, isto é, de base matemática, da relação entre a componente periódica e a componente aperiódica (Yumoto1982) que compõem um segmento sustentado de voz vozeada. A primeira componente decorre da vibração das pregas e a segunda decorre de ruído glótico. A avaliação entre as duas componentes traduz a eficiência do processo de fonação: quanto maior for a eficiência na utilização do fluxo de ar expelido pelos pulmões em energia de vibração das pregas vocais, e quanto mais íntegro (i.e., saudável ou escorreito) for o ciclo vibratório destas pregas, maior será a relação HNR. Inversamente, quanto menor for aquela eficiência ou quanto mais anómalo for o ciclo vibratório, maior será o ruído glótico e mais baixa resultará a

relação HNR. Uma voz saudável deve, assim, caracterizar-se por uma relação HNR elevada, a que se associa a impressão de voz sonora e harmónica. Um baixo HNR denota uma voz asténica e disfónica.

De modo a enquadrar o tratamento matemático das duas componentes fornece-se, na Figura 2, uma possível representação do modelo fonte-filtro para a produção de voz vozeada.

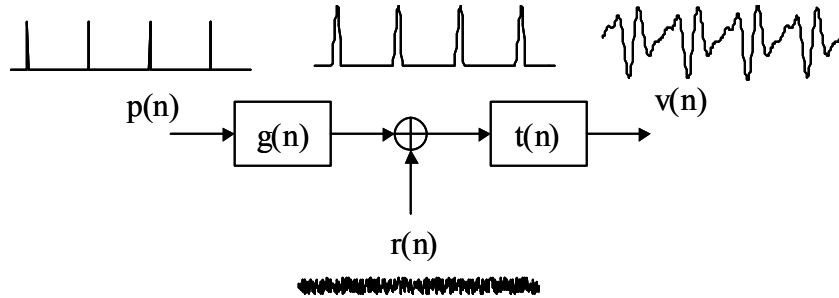


Figura 2: Modelo de produção de fala vozeada. Os impulsos glotais resultam da modelação de uma sequência de impulsos ideais, representados pelo sinal discreto $p(n)$, pelo modelo de impulso glótico $g(n)$. Os impulsos glotais são subsequentemente adicionados a ruído $r(n)$ e o resultado é filtrado por $t(n)$ que modeliza as ressonâncias do tracto vocal. O sinal vozeado é representado por $v(n)$.

A fonte ideal é representada pela sequência de impulsos $p(n)$ de acordo com a Equação (1). Um impulso de índice k tem amplitude a_k e é colocado no instante $kT + \Delta T_k$, em que ΔT denota um pequeno desvio em relação a T . T representa o período fundamental da vibração das pregas vocais e o seu recíproco é $F_0 = 1/T$, também comumente designada de frequência fundamental ou *pitch*. Se a sequência de impulsos for perturbada em amplitude, a_k será distinto para cada valor de k . Por outras palavras, a sequência será afectada de *shimmer*. O *shimmer* será nulo se a_k não depender de k , ou seja, se a_k for constante.

$$p(n) = \sum_k a_k \delta(n - kT - \Delta T_k) \quad (1)$$

Se a sequência de impulsos for perturbada por uma irregularidade na colocação temporal dos impulsos, o valor de ΔT_k será distinto para cada valor de k . Por outras palavras, a sequência será afectada de *jitter*. O *jitter* será nulo se ΔT_k for constante para todos os valores de k . Idealmente, $\Delta T_k = \Delta T = 0$. Na sequência deste artigo, admitimos esta última condição e também que $a_k = 1$ para qualquer valor de k .

De acordo com o modelo fonte-filtro da Figura 2, a sequência ideal de impulsos $p(n)$ é filtrada por $g(n)$ que representa a forma de onda de um único impulso glotal. Esta operação é equivalente à convolução entre $p(n)$ e $g(n)$ e o resultado é uma sequência de impulsos glotais, também ilustrada na Figura 2. Esta sequência é adicionada a ruído estacionário, representado por $r(n)$, e o conjunto é filtrado por $t(n)$ que modeliza as ressonâncias do tracto vocal³ (i.e., corresponde à resposta ao impulso da função de transferência que modeliza o tracto vocal). O resultado, $v(n)$, consiste no sinal de voz. Estas operações são traduzidas pela Equação 2.

³ Incluindo também as ressonâncias próprias da cavidade oral e nasal.

$$v(n) = [p(n) * g(n) + r(n)] * t(n) \quad (2)$$

No domínio de Fourier, ou das frequências, a operação de convolução dá lugar à operação de multiplicação e a Equação (2) tem a forma da Equação (3). $V(\omega)$ representa a transformada de Fourier de $v(n)$ e identicamente para as restantes quantidades (ω representa a variável de frequência e n representa a variável temporal).

$$V(\omega) = [P(\omega) \times G(\omega) + R(\omega)] \times T(\omega) = H(\omega) + N(\omega) \quad (3)$$

A representação no domínio da frequência do sinal de voz, $V(\omega)$, tem assim duas componentes: $H(\omega)$ e $N(\omega)$, tal como clarifica na Equação (4). A primeira traduz a representação espectral dos impulsos glóticos filtrados pelo tracto vocal e a segunda modeliza o ruído glótico filtrado pelo tracto vocal.

$$\begin{aligned} H(\omega) &= P(\omega) \times G(\omega) \times T(\omega) \\ N(\omega) &= R(\omega) \times T(\omega) \end{aligned} \quad (4)$$

A relação HNR é, por definição, uma medida logarítmica da relação das energias associadas às duas componentes, o que presume a integração da potência espectral ao longo da gama audível de frequências:

$$HNR = 10 \times \log_{10} \frac{\int_{\omega} |H(\omega)|^2}{\int_{\omega} |N(\omega)|^2}. \quad (5)$$

O uso da medida logarítmica é pertinente porque estabelece uma boa correlação entre a intensidade física do som e a sua percepção, a que se associa o conceito de *loudness*. Por outras palavras, o HNR tenta medir a relação entre a percepção da componente periódica de um som vozeado, e a percepção da componente de ruído desse sinal.

2.2. Métodos de cálculo

Na prática, o cálculo do espectro é realizado através de técnicas eficientes como a *Fast Fourier Transform* (FFT) que é um método de cálculo rápido da Transformada de Fourier. Deste modo, o espectro é calculado não como uma função contínua, mas como uma amostragem desta função pelo que, na prática, o operador integração dá lugar ao somatório. Em consequência, a Equação (5) apresenta-se da seguinte forma:

$$HNR = 10 \times \log_{10} \frac{\sum_k |H(\omega_k)|^2}{\sum_k |N(\omega_k)|^2}. \quad (6)$$

Há autores (Murphy, 2008) que argumentam que a expressão anterior não é apropriada para avaliar o ruído glótico anterior à interacção com o tracto vocal. Por outras palavras, se se pretender avaliar o HNR da fonte glótica em vez do HNR decorrente do sinal de voz, o efeito do tracto vocal deverá ser cancelado o que, considerando a substituição das

expressões da Equação (4) na Equação (5), conduz ao seguinte cálculo alternativo para o HNR, também designado de GHNR:

$$GHNR = 10 \times \log_{10} \frac{1}{N} \sum_k \frac{|H(\omega_k)|^2}{|N(\omega_k)|^2}, \quad (7)$$

em que N é o número de pontos do somatório. No nosso contexto estamos, contudo, interessados na avaliação da relação harmónicos-ruído que esteja em linha com a avaliação perceptiva a partir do sinal de voz, o que implica a inclusão da influência do tracto vocal. Deste modo, assumimos na sequência deste artigo a definição de HNR de acordo com a Equação (6).

O cálculo do HNR encerra uma dificuldade básica: o sinal captado por um microfone consiste em $v(n)$, cuja transformada de Fourier é $V(\omega)$. As componentes $H(\omega)$ e $N(\omega)$ encontram-se combinadas no sinal, pelo que é necessário separá-las. Os métodos de cálculo propostos por vários autores diferem, sobretudo, nas técnicas de estimação daquelas componentes a partir de $V(\omega)$.

Para melhor se caracterizar o problema e, também, ser possível avaliar o desempenho de soluções alternativas de cálculo do HNR, implementou-se um algoritmo de geração de voz sintética de acordo com o diagrama de blocos da Figura 2. Este algoritmo foi implementado no curso do desenvolvimento de uma aplicação comercial para PC, de apoio à análise e diagnóstico da voz⁴. Deste modo, é possível gerar sinais de voz sintética com valores pré-determinados de *jitter*, *shimmer* e HNR, o que permite quantificar o desempenho de métodos alternativos de extracção de parâmetros de perturbação e qualidade de voz. Por exemplo, configurando a frequência fundamental da voz sintética para $F_0 = 250$ ciclos por segundo (ou Hertz), adoptando um modelo de impulso glotal realista, usando para o tracto vocal o modelo correspondente à vogal /a/, e considerando o ruído $r(n)$ nulo; obteve-se um sinal de voz cujo espectro se ilustra na Figura 3.

⁴ Designada de VoiceStudio. Encontra-se descrita juntamente com uma versão de demonstração em <http://www.seignal.pt>

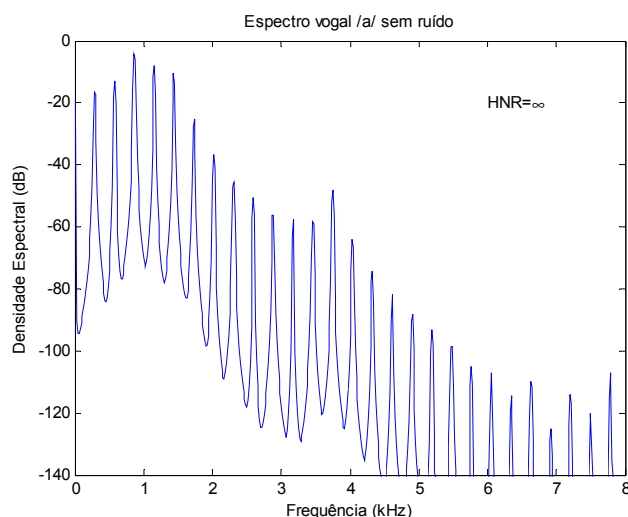


Figura 3: Representação do espectro de voz sintética correspondente à vogal /a/ na ausência de ruído glótico. A estrutura espectral do sinal é escorreta, com uma boa presença e regularidade harmónica.

Esta figura revela que a regularidade harmónica do espectro é perfeita, sem descontinuidades ou perturbação por ruído. Se, porém, se adicionar ruído configurado para que a relação HNR resultante seja 20 dB, obtém-se o espectro representado na Figura 4.

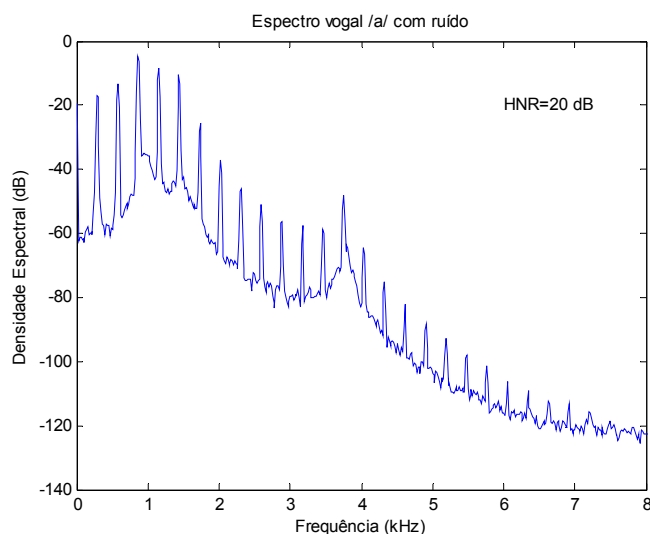


Figura 4: Representação do espectro de voz sintética correspondente à vogal /a/ afectada por ruído glótico. A influência do ruído é visível pela ‘contaminação’ que introduz nos harmónicos do sinal, podendo alguns ficar, inclusivamente, ‘apagados’ pelo ruído.

Esta figura permite clarificar que, em relação à Figura 3, o ruído diminui a pureza da estrutura harmónica, reduzindo muito o destaque de cada harmónico (ou parcial da estrutura harmónica) em relação ao ruído, podendo inclusivamente ‘apagar’ a sua presença.

Se no algoritmo de produção de voz sintética se suprimir a geração de impulsos glóticos mantendo só o ruído, obtém-se o espectro representado na Figura 5.

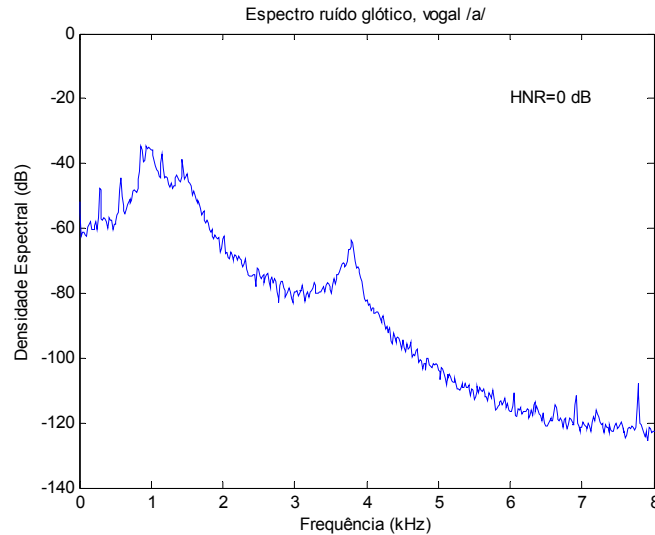


Figura 5: Representação do espectro de voz sintética correspondente à vogal /a/ quando a fonte é constituída só por ruído glótico.

Como seria de esperar, verifica-se neste caso que para além de alguns picos espúrios, não há evidência na representação espectral de uma estrutura harmónica.

Estas figuras permitem também compreender melhor o desafio no cálculo do HNR: ao captar um sinal de voz, o espectro a que se tem acesso é o ilustrado na Figura 4, ou seja, a $V(\omega)$. Como base neste, dever-se-á usar uma estratégia para obter as componentes ilustradas nas Figuras 3 e 5 de modo a ser viável calcular a sua potência espectral e, a partir desta, a relação HNR.

2.2.1. Método utilizado no Praat

Um *software* muito popular no meio académico e de investigação é o Praat⁵. O Praat utiliza para o cálculo do HNR um algoritmo desenvolvido por Boersma (Boersma 1993). A sua abordagem é indirecta já que não realiza a separação de componentes atrás descrita, seguindo antes um procedimento baseado nas propriedades da função auto-correlação. De facto, calculando a autocorrelação (AC) do sinal de voz $v(n)$, obtém-se:

$$AC_V(\tau) = \sum_n v(n) \times v(n + \tau) = AC_H(\tau) + 2 \times CC_{H,N}(\tau) + AC_N(\tau), \quad (8)$$

em que o operador CC representa a correlação cruzada e os índices H e N simbolizam, respectivamente, as componentes harmónica e ruído do sinal de voz. Dado que as componentes harmónica e de ruído se consideram independentes e, portanto, não-correlacionadas, a sua correlação cruzada é nula pelo que a equação anterior se reduz a

$$AC_V(\tau) = AC_H(\tau) + AC_N(\tau). \quad (9)$$

⁵ <http://www.praat.org/>

Por definição, quando o parâmetro τ é nulo, a função $AC_V(\tau)$ exibe um máximo global que traduz a potência total do sinal de voz que, por sua vez, resulta da soma de potência das componentes harmónica e de ruído:

$$AC_V(0) = AC_H(0) + AC_N(0). \quad (10)$$

Admitindo que o ruído é branco (ou de densidade espectral plana), a função $AC_N(\tau)$ é nula para $\tau \neq 0$. Por outro lado, admitindo estacionaridade, a função $AC_H(\tau)$ é periódica e, em particular, o valor de $AC_H(0)$ repete-se quando τ é múltiplo inteiro do período fundamental da voz, $T=1/F_0$. Por outras palavras, a função autocorrelação do sinal de voz (vozeada) exibe máximos locais para valores de τ múltiplos inteiros do período fundamental. Assim, para encontrar a relação HNR basta calcular a função autocorrelação do sinal de voz, identificar o primeiro máximo local e ler o valor correspondente à potência da componente harmónica pois $AC_V(T)=AC_H(T)=AC_H(0)$. O valor de potência correspondente à componente de ruído determina-se usando a Equação (10). O valor HNR calcula-se, finalmente, através de:

$$HNR = 10 \times \log_{10} \frac{AC_V(T)}{AC_V(0) - AC_V(T)}. \quad (11)$$

Na prática, o cálculo do HNR por esta via indirecta exige alguns cuidados especiais, além de que os pressupostos de estacionaridade e ruído branco, em rigor, não se verificam.

2.2.2. Método utilizado no Dr. Speech

O Dr. Speech é um outro *software*⁶ de avaliação acústica frequentemente usado em meio clínico ou académico (Mendes 2006). Este *software* utiliza o algoritmo de Yumoto e Gould (Yumoto 1982) cuja técnica se fundamenta num facto simples: dado que a componente de ruído de um sinal de voz pode ser modelizada como ruído aditivo de média nula, a componente harmónica, $v_H(t)$, pode ser estimada somando um número elevado (K) de períodos do sinal, após alinhamento cuidadoso:

$$v_H(t) = \frac{1}{K} \sum_k v(t - T_k), \quad 0 \leq t \leq T. \quad (12)$$

A necessidade de alinhamento decorre da circunstância da voz natural exibir sempre algum nível de *jitter*. Por esta razão, na Equação (12) ter-se-á que $T_k \neq kT$, caso contrário o *jitter* seria nulo e o sinal seria perfeitamente periódico. A componente de ruído pode ser estimada calculando a diferença entre cada período do sinal de voz e o sinal $v_H(t)$, após alinhamento. Deste modo, a relação HNR é obtida calculando:

⁶ <http://www.drspeech.com>

$$HNR = 10 \times \log_{10} \frac{K \int_0^T |v_H(t)|^2 dt}{\sum_k \int_0^T |v(t - T_k) - v_H(t)|^2 dt}. \quad (13)$$

Sendo uma técnica do domínio do tempo, possui a vantagem de ser computacionalmente simples mas encerra algumas fragilidades, como por exemplo, ser muito sensível a desalinhamentos e não admitir ruído de natureza não-linear. Os autores reconhecem inclusivamente que o método possa ser inválido em casos de disфонia.

2.2.3. Outros métodos

Uma outra abordagem de cálculo do HNR foi inicialmente proposta por Krom (Krom 1993) e subsequentemente modificada por Qi (Qi 1997). Esta abordagem baseia-se na propriedade do cepstrum⁷ permitir desacoplar a componentes de variação rápida do espectro (relacionadas com os harmónicos) e as componentes de variação lenta do espectro (relacionadas com a envolvente espectral que retrata razoavelmente o perfil do ruído e, portanto, os formantes). Deste modo -identificando os picos do espectro correspondentes às componentes harmónicas e usando diversos passos de filtragem, que permitem obter uma estimativa do espectro do ruído- é possível calcular o HNR através da Equação (6). Apesar de mais directa, esta abordagem é vulnerável à natureza dos sinais de voz e, em particular, os seus resultados dependem muito da frequência fundamental F_0 . Estes problemas foram subsequentemente minimizados em novos resultados publicados por Murphy (Murphy 2007).

2.3. Novo procedimento de cálculo

Foi desenvolvido um novo método de cálculo que procura atender à definição da medida HNR, através da segmentação das componentes harmónica e ruído a partir do sinal de voz captado por um microfone. Contrariamente aos métodos anteriormente referidos que efectuem uma análise de sinal com o objectivo de estimar a potência das duas componentes, o novo método combina funções de análise e síntese de sinal com o objectivo de segmentar fisicamente as duas componentes e permitir a sua reconstrução individual. Esta é uma capacidade de processamento de sinal inovadora e inspira-se em técnicas de estimação precisa de componentes sinusoidais no sinal de voz (Ferreira 2001a, Ferreira 2005), e em técnicas de reconstrução no tempo de sinais a partir de modelos no domínio da frequência (Ferreira 2001b).

O procedimento matemático subjacente pode ser sintetizado nos seguintes passos principais de processamento de sinal:

⁷ O cepstrum aqui considerado (cepstrum real) consiste na transformada de Fourier inversa do logaritmo do espectro (o que explica a designação de ‘ceps’ como inverso de ‘spec’). Remete portanto para um domínio do tempo que caracteriza a periodicidade existente no espectro. Em termos práticos, é útil por exemplo para calcular o período fundamental (sem segundos) de uma estrutura harmónica.

- é efectuada uma análise precisa de todas as componentes sinusoidais existentes no sinal de voz e, a partir destas, é identificada a estrutura harmónica do sinal de voz mais plausível,
- a estrutura harmónica mais plausível é modelizada parametricamente e é reconstruída no domínio das frequências,
- a estrutura harmónica reconstruída no domínio das frequências é subtraída ao espectro complexo do sinal de voz (isto é, incluindo a informação quer de magnitude, quer de fase) de forma a obter-se um resíduo espectral que corresponde à estimativa do espectro do ruído,
- as representações espectrais da estrutura harmónica e da estimativa do ruído são usadas quer para determinar a sua potência média, quer para reconstruir os correspondentes sinais no domínio do tempo. As duas medidas de potência assim obtidas são usadas para calcular a relação HNR de acordo com a Equação (6).

Este procedimento tem duas vantagens. Por um lado é um método directo de análise precisa, segmentação e síntese das componentes harmónica e ruído, o que encerra o potencial de permitir medições mais rigorosas. Por outro lado, permite a reconstrução isolada das componentes harmónica e de ruído para o domínio do tempo, o que viabiliza a sua audição e gravação individual. Isto corresponde a obter os sinais das Figuras 5 e 3 a partir do sinal da Figura 4. Esta capacidade é inédita em *softwares* de análise acústica e encontra-se disponível no *software* VoiceStudio de apoio à análise e diagnóstico da voz. A Figura 6 ilustra o ambiente do VoiceStudio e o relatório de parâmetros acústicos.

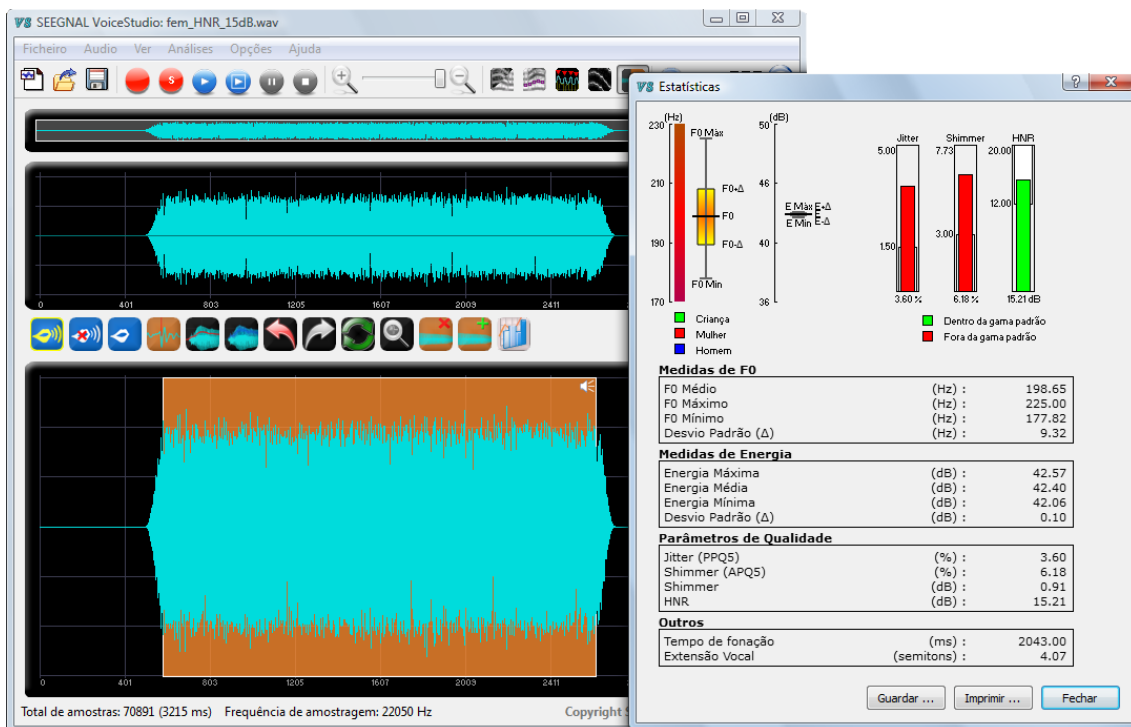


Figura 6: Ilustração do ambiente do VoiceStudio e do relatório de parâmetros acústicos relativos à região do sinal de voz destacada.

3. Desempenho no cálculo do HNR

3.1. Usando voz sintética

O novo procedimento de cálculo do HNR, disponível no *software* VoiceStudio, foi avaliado utilizando o algoritmo de síntese de voz sintética já referido na secção 2.2. Utilizou-se um modelo da vogal /a/ para a síntese de voz e usaram-se duas frequências fundamentais, $F_0=100$ Hz e $F_0=200$ Hz, de modo a simular uma voz masculina e outra feminina. Para cada género geraram-se cinco versões da vogal /a/ com valores de HNR predefinidos e correspondentes a 5 dB, 10 dB, 15 dB, 20 dB e 25 dB. Incluem-se, assim, gamas de HNR associadas a valores patológicos e a valores normais (superiores a cerca de 12 dB). Para referência, usaram-se dois *softwares* conhecidos para efectuar as mesmas medições de HNR: Dr. Speech e Praat.

As Figuras 7 e 8 apresentam os resultados para as vozes masculina e feminina, respectivamente.

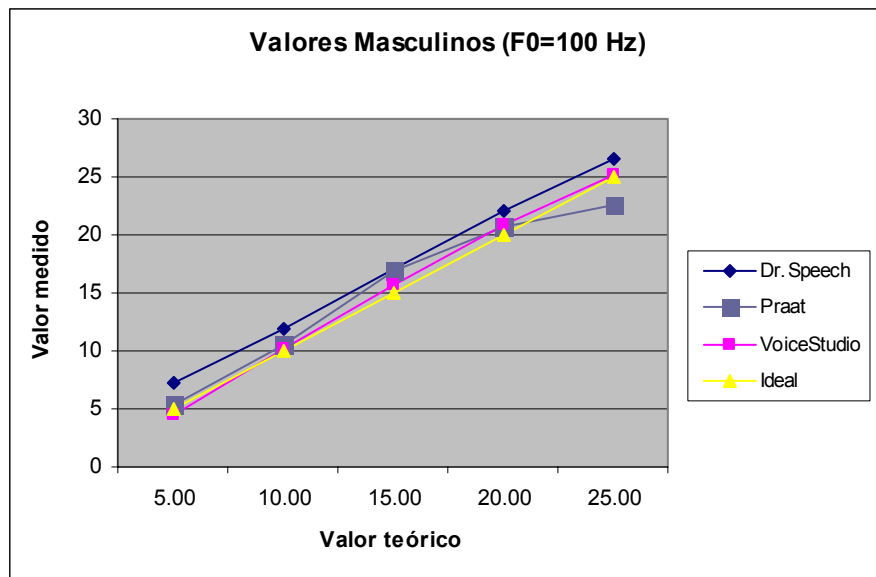


Figura 7: Resultados da medição do HNR pelos *softwares* Dr. Speech, Praat e VoiceStudio, de vários sinais sintéticos de voz (vogal /a/) com $F_0=100$ Hz e valores de HNR predeterminados entre 5 e 25 dB .

A Figura 7 permite concluir que em relação à curva ideal (assinalada por triângulos), os resultados fornecidos pelo Dr. Speech são os que mais divergem, enquanto que os valores obtidos através do Praat e VoiceStudio são mais concordantes com os teóricos, com uma pequena vantagem para este último.

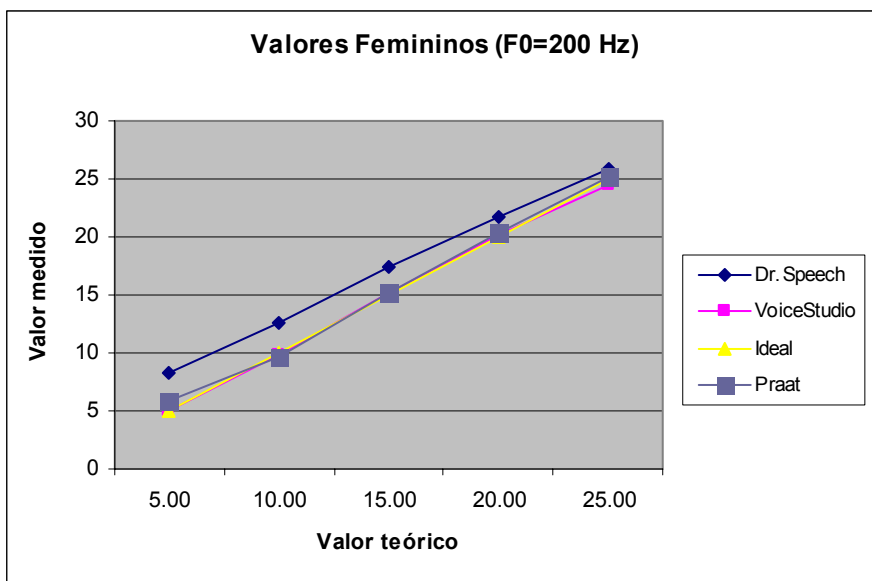


Figura 8: Resultados da medição do HNR pelos *softwares* Dr. Speech, Praat e VoiceStudio, de vários sinais sintéticos de voz (vogal /a/) com $F_0=200$ Hz e valores de HNR predeterminados entre 5 e 25 dB .

A Figura 8 permite concluir que em relação à curva ideal (assinalada por triângulos), os resultados fornecidos pelo Praat e VoiceStudio praticamente coincidem com os valores teóricos, enquanto que os fornecidos pelo Dr. Speech exibem um desvio significativo e sistemático, apesar de diminuir para valores teóricos mais elevados.

Os resultados das duas figuras anteriores foram combinados para exprimir o desvio médio dos valores fornecidos por cada *software*, em relação ao ideal. Esta avaliação média encontra-se representada na Figura 9.

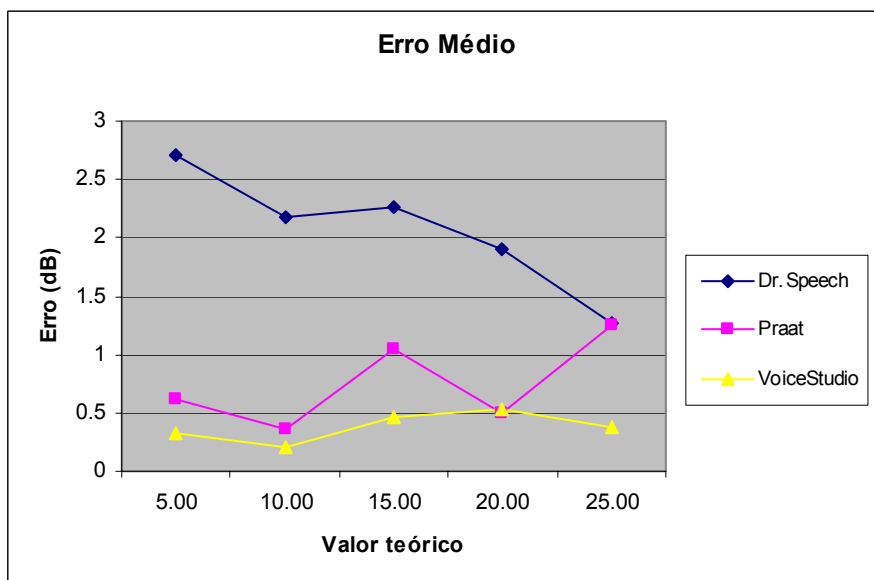


Figura 9: Erro médio na medição do HNR pelos *softwares* Dr. Speech, Praat e VoiceStudio.

Esta figura permite concluir que o desvio mais acentuado se verifica, de modo sistemático, nas medições de HNR pelo Dr. Speech, podendo atingir cerca de 2.5 dB. Este desvio é muito considerável e pouco abonatório para a precisão do procedimento de cálculo do HNR pelo Dr. Speech. Por outro lado, os *softwares* Praat e Voice Studio são os que apresentam desvios de HNR mais contidos e uniformes para a gama de valores considerada. Dado que, em média, o desvio é, em ambos os casos, inferior a 1 dB, pode-se afirmar que a precisão dos respectivos algoritmos de cálculo de HNR é aceitável para efeitos práticos, havendo contudo uma ligeira vantagem para os resultados do VoiceStudio no conjunto de testes considerado. De modo a permitir avaliações independentes, todos os ficheiros de áudio usados neste teste de desempenho encontram-se disponíveis na versão de demonstração do VoiceStudio⁸.

3.2. Usando voz natural

Actualmente a avaliação da voz natural e das suas perturbações (disfonia) recorre a diversos métodos de caracterização, mais objectivos ou subjectivos. Saliente-se a avaliação perceptiva por um ouvinte experiente, as medidas acústicas, as mensurações fisiológicas e, também, a auto-avaliação do impacto da perturbação vocal na qualidade de vida do falante (Bhutta *et al.*, 2004).

A análise acústica da voz patológica assume várias vantagens, como sejam a sua mensuração quantitativa, o facto de consistir num procedimento não invasivo e, também, a sua eficiência em termos de custos e tempo. Como desvantagem aponte-se que a maioria das análises acústicas se baseia em ondas quase-periódicas, pelo que o seu uso é difícil na análise de vozes com muito ruído ou irregulares (Martens *et al.* 2007).

Na década de 90 começaram a ser publicados os primeiros estudos sobre a análise e interpretação das características da voz humana através dos diferentes parâmetros acústicos que compõem o sinal: periodicidade, amplitude, duração e composição espectral. Assim, investigadores concluíram da relação entre *loudness* e amplitude (Grenn 1993), assim como entre *pitch* e frequência (Moore 1993).

A identificação do ruído no sinal de voz é considerado por muitos como essencial para determinar e quantificar as características acústicas da voz disfónica. Isto porque quanto maior a componente de ruído, mais este se manifesta na diminuição da regularidade vibratória da ondulação da mucosa das pregas vocais (Dekrom *et al.* 1995). Note-se que um valor baixo de HNR é encontrado em 83% das vozes patológicas (Parsa 2000).

Como referido na secção 2.1, o HNR traduz a incidência de ruído glótico e, portanto, exprime as suas perturbações em amplitude e frequência. A magnitude do ruído glótico corresponde, de forma marcada, a uma qualidade vocal soprosa, assim como à percepção de rouquidão (Huang *et al.* 1997).

Este trabalho baseou-se na análise e classificação de uma base de dados de vozes patológicas⁹ compostas por 12 casos de falantes femininos e 11 masculinos. Em cada sub-grupo existem exemplos de três situações passíveis de causar perturbações vocais: fendas glóticas (por alterações da mucosa das pregas vocais ou de causa neurológica),

⁸ Descarregável a partir do endereço Web: <http://www.seegnal.pt>

⁹ Que acompanha a versão comercial do VoiceStudio.

lesões de massa (benignas, malignas ou iaterogénicas) e, ainda, patologias sub e supra-glóticas com reflexo no controle pneumofónico (linfomas e tumores). Embora significativa em termos de variedade de patologias representadas – em cada um dos géneros estudados – é uma amostra com necessidade de aumento da casuística. Acrescente-se, ainda, que vai de encontro à distribuição que estes casos apresentam numa consulta hospitalar de Otorrinolaringologia e Terapia da Fala.

Todas as amostras de voz foram analisadas quanto ao HNR por três *softwares* de análise acústica: VoiceStudio, Dr. Speech e Praat. Os resultados obtidos foram os seguintes (ver Tabela 1):

	VoiceStudio	Praat	Dr. Speech
FEM_01	11.5	14.2	14.5
FEM_02	7.1	8.2	10.2
FEM_03	12.4	17.5	13.7
FEM_04	9.9	12.0	13.9
FEM_05	15.0	19.8	19.3
FEM_06	0.0	0.0	0.0
FEM_07	10.0	16.9	18.8
FEM_08	17.9	21.6	21.8
FEM_09	0.0	10.6	0.0
FEM_10	0.0	9.1	5.8
FEM_11	0.0	4.6	4.8
FEM_12	10.3	13.9	15.5
MAL_01	12.6	13.4	14.4
MAL_02	9.6	10.2	12.7
MAL_03	4.8	9.5	0.0
MAL_04	11.3	13.3	14.4
MAL_05	0.0	9.7	11.6
MAL_06	0.0	0.0	0.0
MAL_07	13.3	15.2	17.0
MAL_08	15.6	17.8	18.4
MAL_09	3.2	13.0	0.0
MAL_10	0.0	2.3	0.0
MAL_11	14.0	16.1	17.8

Tabela 1: Valores de HNR (em dB) obtidos por cada um dos *softwares* de análise acústica usados.

Para as amostras de vozes patológicas analisadas notaram-se as seguintes características:

- O valor mais baixo de HNR foi, sistematicamente, encontrado no sinal vocal do caso com nódulos bilaterais das pregas vocais. Esta é uma patologia estrutural que implica uma lesão de massa, geralmente associada à presença de uma fenda glótica pela existência dos nódulos. Assim, gera-se maior ruído à vibração devido à lesão de massa que interfere com o turbilhão de ar normal.

- Dois dos programas também caracterizaram como reduzido o valor de HNR nos casos de:

- Refluxo gastro-esofágico: esta patologia pressupõe a subida de sucos ou vapores do estômago para a região faringo-laríngea, causando irritação das pregas vocais (edema e rubor) e a formação de granulomas de contacto no terço posterior das pregas vocais. Estas alterações manifestam-se através de uma maior irregularidade de vibração da mucosa e, também, em falhas no fechamento glótico.

- Tumor sub-glótico: a presença de uma massa num nível inferior ao da glote (espaço entre as pregas vocais) provoca, também, um maior descontrolo à passagem do ar pulmonar e a assimetria de vibração laríngea.

Por fim, realizou-se o cálculo da correlação entre os valores de HNR obtidos nos três *softwares* e observaram-se os seguintes resultados (ver Tabela 2):

	VoiceStudio-Dr.Speech	VoiceStudio-Praat	Praat-Dr.Speech
correlação	0.89	0.85	0.82

Tabela 2: Correlação entre os valores de HNR obtidos nas amostras de vozes patológicas entre os três programas de análise acústica estudados.

Note-se que embora todos os programas possuam um algoritmo para cálculo do HNR diferente, obtiveram-se correlações fortes entre as suas medições, especialmente entre as do VoiceStudio e as do Dr. Speech (0,889).

3.3. Correlação do HNR com avaliação perceptiva

Embora a análise acústica seja útil para a quantificação das características e/ou respostas ao tratamento em disfonias, a avaliação perceptiva – realizada por profissionais experientes – assume-se como um método importante e eficiente na caracterização de uma qualquer perturbação vocal (Canitto *et al.*, 2004). Porém, a relação entre estas duas dimensões não é nem óbvia nem directa – são vários os estudos que demonstram inconsistências e acabam por contradizer os resultados obtidos (Dejonckere 1996; Kreiman *et al.*, 1998; Morsomme 2001; Bhuta 2004).

Como a voz é multidimensional, a sua avaliação deve incluir as vertentes objectivas e subjectivas e, por isso, nenhuma medida isolada consegue representar todos os seus aspectos (Bhuta 2004; Eadie *et al.* 2005; Ma & Yiu 2006). A *Japanese Society of Logopedics and Phoniatrics* e a *European Research Group* recomendam a escala perceptiva GRBAS para uso clínico e de investigação. Como indicado na secção 1, esta foi traduzida para o Português por Pinho (2002) com a designação de RASAT (Rouquidão, Aspereza, Soprosidade, Astenia e Tensão).

A revisão da literatura demonstra que a avaliação perceptiva faz depender a sua validade de vários factores, como sejam: o tipo de escala usada, a qualidade vocal e as amostras de voz em análise, a preparação e experiência prévias do avaliador, e a existência de parâmetros vocais externos (ex: fenómenos de co-articulação, características supra-segmentais) que funcionem como ajudas ao ouvinte. Estudos mostram que a variabilidade de classificações de vozes individuais é maior para as ligeira-moderadamente alteradas, do que as dos extremos (normais ou severamente perturbadas) (Yu *et al.* 2001; Eadie *et al.* 2005; Ma & Yiu 2006).

As análises acústicas e perceptivas devem ser entendidas como complementares e estar integradas na avaliação multidimensional da voz disfónica. Esta conclusão é crítica do ponto de vista clínico já que ao assumir esta afirmação se condiciona não só a consistência da avaliação vocal, como também os resultados dos tratamentos. Note-se que nos diferentes estudos de correlação entre as medidas subjectivas e instrumentais a

percentagem de concordância pode variar entre 49,9% (Wuyts *et al.* 2000) e 86,0% (Yu *et al.* 2001).

A classificação perceptiva das amostras de voz estudadas – 12 femininas e 11 masculinas – foi realizada por um Terapeuta da Fala com mais de 8 anos de experiência profissional nesta área clínica.

Assim, neste estudo – nos três *softwares* de análise acústica – correlacionaram-se os valores no HNR resultantes da análise da vogal /a/ sustentada com as classificações perceptivas da escala RASAT, para todas as amostras de voz (ver Tabela 3).

VS-Rouquidão	0.06	DrSpeech-Rouquidão	-0.01	Praat-Rouquidão	0.18
VS-Aspereza	-0.33	DrSpeech-Aspereza	-0.39	Praat-Aspereza	-0.32
VS-Soprosidade	-0.54	DrSpeech-Soprosidade	-0.60	Praat-Soprosidade	-0.70
VS-Astenia	-0.60	DrSpeech-Astenia	-0.65	Praat-Astenia	-0.57
VS-Tensão	-0.27	DrSpeech-Tensão	-0.18	Praat-Tensão	-0.14

Tabela 3: Correlação entre os valores de HNR e os parâmetros perceptivos, para os três programas de análise acústica estudados (VoiceStudio –VS, Dr. Speech e Praat).

A análise da tabela permite concluir da não existência de associação entre o HNR e o parâmetro “rouquidão”. A “aspereza” obteve valores muito fracos de correlação. Este é classificado de acordo com a percepção de ruídos adventícios, em especial nas altas frequências, e instabilidade de fonação gerada, principalmente por padrões de hiperfuncionamento (ou hiperfunção) laríngea. O valor de “soprosidade” equivale, na prática clínica, à perturbação do encerramento glótico (ou fenda). Do mesmo modo, a existência de alterações na aproximação das pregas vocais origina uma maior turbulência à passagem do ar pulmonar que, acusticamente, corresponde a ruído. Por outro lado, este descontrole da coluna de ar pulmonar implica uma menor eficácia pneumofônica que gera cansaço vocal – “astenia” – mais rápida e notada pelo falante. O parâmetro “tensão” também não se correlacionou de modo significativo com o HNR.

Globalmente, pode concluir-se que para a base de dados considerada existe uma correlação significativa e consistente para os *softwares* usados, entre o parâmetro acústico HNR e os parâmetros perceptivos ‘soprosidade’ e ‘astenia’. Observa-se também que os valores de correlação entre a análise acústica e perceptiva, para os três *softwares*, se inserem no intervalo encontrado pelas investigações revistas, sendo que a correlação mais forte entre o valor de HNR e o parâmetro “soprosidade” foi apresentada pelo programa Praat, enquanto que em relação ao parâmetro “astenia” o programa que denotou maior correlação acústico-perceptiva foi o Dr. Speech.

4. Conclusão

Neste artigo caracterizou-se o parâmetro acústico Harmonics-to-Noise Ratio (HNR) quanto à sua natureza, sua pertinência no diagnóstico da voz e métodos alternativos para o seu cálculo. Introduziu-se também uma nova abordagem de cálculo cujo desempenho foi avaliado usando registos de voz sintética com valores pré-definidos de HNR e tendo por referência outros algoritmos. A correlação entre análise acústica e análise perceptiva foi também estudada com base numa amostra de 23 vozes patológicas, tendo-se

concluído que existe uma correlação significativa, para as amostras consideradas, entre os valores de HNR e as avaliações por um profissional da terapia da fala, dos parâmetros ‘soprosidade’ e ‘astenia’ da escala RASAT. Este trabalho sugere novos estudos mais abrangentes e aprofundados da relação entre parâmetros acústicos e parâmetros perceptivos, com base numa amostra mais ampla e representativa da variedade vocal ao longo do ciclo vital, nos dois géneros e de acordo com diferentes etiologias e patologias laríngeas.

5. Referências

(Bhuta 2004) Bhuta, T.; Patrick, L.; Garnett, J.D. *Perceptual Evaluation of voice quality and its correlation with acoustic measurements*. Em *Journal of Voice*. Vol. 18, páginas 299-304, 2004.

(Boersma 1993) Paul Boersma. *Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound*. Em *Proceedings of the Institute of Phonetic Sciences*, Vol. 17, páginas 97-110, 1993.

(Canitto *et al.*, 2004) Canitto, M.P.; Woodson, G.E.; Murry, T.; Bender, B.. *Perceptual Analyses of Spasmodic Dysphonia Before and After Treatment*. Em *Archive Otolaryngology Head and Neck Surgery*, Vol. 130, páginas 1393-1399, 2004.

(Dekrom 1995) Deckrom, G. *Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments*. *Journal of Speech and Hearing Research*, Vol. 38, páginas 794-811, 1995.

(Dejonckere 1996) Dejonckere P.H.; Remacle, M.; Fresnel-Elbaz E.; Woisard, V.; Crevier-Buchman, L.; Millet, B. *Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements*. *Revue de Laryngologie, Otologie and Rhinologie*. Vol. 117, páginas 219-224, 1996.

(Eadie *et al.* 2005) Eadie, T.; Doyle, F.. *Classification of Dysphonic Voice: Acoustic and Auditory-Perceptual Measures*. *Journal of Voice*. Vol. 19, páginas 1-14, 2005.

(Ferreira 2001a) Aníbal Ferreira. *Accurate Estimation in the ODFT Domain of the Frequency, Phase and Magnitude of Stationary Sinusoids*. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, páginas 47-50, New Paltz, E.U.A., Outubro de 2001.

(Ferreira 2001b) Aníbal Ferreira. *Combined Spectral Envelope Normalization and Subtraction of Sinusoidal Components in the ODFT and MDCT Frequency Domains*. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, páginas 51-54, New Paltz, U.S.A, Outubro de 2001.

(Ferreira 2005) Aníbal Ferreira e Deepen Sinha. *Accurate and Robust Frequency Estimation in the ODFT Domain*. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, E.U.A., Outubro de 2005.

(Guimarães 2007) Isabel Guimarães. *A Ciência e a Arte da Voz Humana*. Escola Superior de Saúde do Alcoitão (ESSA), Alcabideche, 2007.

(Green 1993) Green, D.M. *Auditory Intensity Discrimination*. Em: Yost, W.A.; Pooper, A.N.; Fay, R.R. Edit. *Human Psychophysics*. New York: Springer; 1993: 13-55.

(Huang *et al.* 1997) Huang, D.Z.; Lin, S.; O'Brien, R. *Dr. Speech User's Guide*. s.l.: Tiger Electronics Inc. 1997.

(Kreiman *et al.* 1998) Kreiman, J.; Gerratt, B.R. *Validity of rating scale measures of voice quality*. Journal Acoustic Society America, Vol. 104, páginas 1598-1608, 1998.

(Ma & Yiu 2006) Ma, E.P-M.; Yiu, E.M-L. *Multiparametric Evaluation of Dysphonic Severity*. Journal of Voice. Vol. 20, páginas 380-390, 2006.

(Martens 2007) Martens, J.; Versnel, H.; Dejonchere, P. *The Effect of Visible Speech in the Perceptual Rating of Pathological Voices*. Archives of Otolaryngology Head and Neck Surgery, Vo. 133, Páginas 178-185, 2007.

(Mendes 2006) Ana Mendes. *Voice acoustic patterns of patients diagnosed with vibroacoustic disease*. Revista Portuguesa de Pneumologia, vol. XII, n.º 4, Julho/Agosto 2006.

(Moore 1993) Moore, B.C.J.. *Frequency analysis and pitch perception*. Em: Yost, W.A.; Pooper, A.N.; Fay, R.R. Edit. *Human Psychophysics*. New York: Springer; 1993: 56-115.

(Morsomme 2001) Morsomme, D.; Jamart, J.; Werry, C. ; Giovanni, A. ; Remacle, M. *Comparison between the GIRGAS scale and the acoustic and aerodynamic measures provided by EVA for the assessment of dysphonia following vocal fold paralysis*. Folia Phoniatria, Logopedica, Vol. 53, Páginas 317-325, 2001.

(Murphy 2007) Peter J. Murphy e Olatunji O. Akande. *Noise estimation in voice signals using short-term cepstral analysis*. Journal of the Acoustic Society of America, Vol. 121, n.º 3, páginas 1679-1689, Março de 2007.

(Murphy 2008) Peter J. Murphy, Kevin G. McGuigan, Michael Walsh e Michael Colreavy. *Investigation of a glottal related harmonics-to-noise ration and spectral tilt as indicators or glottal noise in synthesized and human voice signals*. Journal of the Acoustic Society of America, Vol. 123, n.º 3, páginas 1642-1652, Março de 2008.

(Parsa *et al.* 2000) Parsa, V.; Jamieson, D.G. *Identification of pathological voices using glottal measures*. Journal of Speech Hearing Research, Vol. 43, Páginas 469-485, 2000.

(Pinho 2002) Pinho, S. *Escala de Avaliação perceptiva da fonte glótica: RASAT*. Voxbrasilis. Vol. 3, páginas 11-13. 2002.

(Qi 1997) Yinyoung Qi e Robert E. Hillman. *Temporal and spectral estimations of harmonics-to-noise ratio in voice signals*. Journal of the Acoustic Society of America, Vol. 102, n.º 1, páginas 537-543, Julho de 1997.

(Krom 1993) G. Krom. *A cepstrum based technique for determining a harmonics-to-noise ratio in speech signals*. Journal of the Hearing Research, Vol. 36, páginas 254-266, 1993.

(Wuyts *et al.* 2000) Wuyts, F.L.; De Bodt, M.S. ; Molenberghs, G.; Remacle, M. Heyler, L.; Millett, B. *The Dysphonic Severity Index: an objective measure of vocal quality based on a multiparameter approach*. Journal of Speech, Language and Hearing Research. Vol. 43, páginas 796-809, 2000.

(Yumoto 1982) E. Yumoto. *Harmonic-to-noise ratio as index of a degree of hoarseness* Journal of the Acoustic Society of America, Vol. 71, n° 6, páginas 1544-1689, Junho de 1982.

(Yu *et al.* 2001) Yu, P.; Ouaknine, M. ; Giovanni, A. *Objective voice analysis for dysphonic patients : a multiparametric protocol including acoustic and aerodynamic measurements*. Journal of Voice. Vol. 15, páginas 529-542, 2001.