República Federativa do Brasil
Ministério do Desenvolvimento, Indústria e do Comércio Exterior
Instituto Nacional da Propriedade Industrial.

(21) **PI0707343-7 A2**

BRPI0707343A2

(22) Data de Depósito: 30/01/2007
(43) Data da Publicação: 03/05/2011
(RPI 2104)

(51) Int.Cl.:
H04M 3/22
H04B 3/46
G10L 19/00

(54) Título: **MÉTODO E APARELHO DE AVALIAÇÃO DE QUALIDADE DE SINAL NÃO INTRUSIVO**

(30) Prioridade Unionista: 31/01/2006 US 60/763383

(73) Titular(es): Tlefonaktiebolaget LM Ericson (publ)

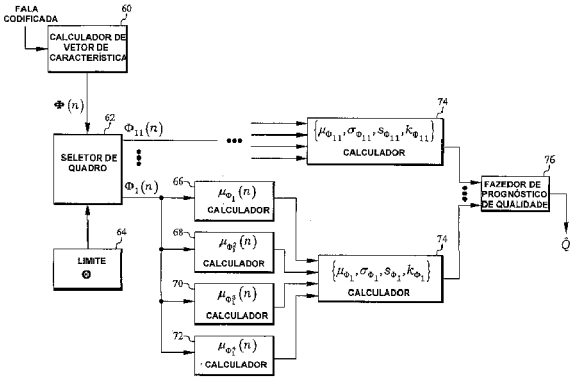
(72) Inventor(es): Bastian kleijn, Stefan Bruhn, Volodya Grancharov

(74) Procurador(es): Momsen, Leonardos & CIA.

(86) Pedido Internacional: PCT SE2007000080 de 30/01/2007

(87) Publicação Internacional: WO 2007/089189de 09/08/2007

(57) Resumo: MÉTODO E APARELHO DE AVALIAÇÃO DE QUALIDADE DE SINAL NÃO INTRUSIVO. Um aparelho de avaliação de qualidade de sinal não intrusivo nclui um calculador de vetor de característica (60) que determina parâmetros epresentando quadros de um sinal e extrai uma coleção de vetores de característica por quadro (Φ(n)) representando informação estrutural do sinal a partir dos parâmetros. Um seletor de quadro (62) preferencialmente eleciona somente quadros (Ω) com um vetor de característica (Φ (n)) ficando entro de uma janela de múltiplas dimensões predeterminada (θ) . Meios (66, 68, 70, 72, 74) determinam um conjunto de características global (Ψ) sobre a oleção de vetores de características (Φ (n)) de momentos estatísticos dos omponentes de vetor de característica selecionados ((1^,02,...O11) . Um previsor de qualidade (76) faz prognóstico de uma medida de qualidade de sinal (QJ do conjunto de características global (Ψ).



"MÉTODO E APARELHO DE AVALIAÇÃO DE QUALIDADE DE SINAL NÃO INTRUSIVO"

CAMPO TÉCNICO

A presente invenção se refere a avaliação de qualidade de sinal não intrusiva e, de forma especial à avaliação de qualidade de fala não intrusiva.

CONHECIMENTO

Avaliação de qualidade de fala é um problema importante em comunicações móveis. A qualidade de um sinal de fala é uma medida subjetiva. esta pode ser expressa em termos de como natural os sons do sinal ou quanto esforço é exigido para entender a mensagem. Em um teste subjetivo, fala é executada para um grupo de ouvintes, os quais são solicitados a avaliar a qualidade deste sinal de fala, ver [1], [2].

A medida mais comum para opinião do usuário é o graduação de opinião média (MOS), obtido através da ponderação de classificações de categoria absolutas (ACR). Em ACR, ouvintes comparam o sinal distorcido com seu modelo interno de fala de alta qualidade. Em testes de degradação de MOS (DMOS), os sujeitos primeiro ouvem a fala original, e então são solicitados a selecionar a classificação de categoria de degradação (DCR) correspondente para a distorção do sinal processado. Testes de DMOS são mais comum em avaliação de qualidade de áudio qualidade, ver [3], [4].

Avaliação da qualidade ouvida como descrita em [l]-[4] não é a somente forma de monitoração de qualidade of serviço (QoS). Em muitos casos, testes subjetivos conversacionais, ver [2], são os métodos preferidos de avaliação subjetiva, onde participantes mantêm conversação sobre um número de redes diferente e votam na sua percepção de qualidade conversacional. Um modelo objetivo de qualidade conversacional pode ser encontrado em [5]. Ainda uma outra classe de monitoração de QoS consiste de testes de inteligibilidade. Os testes de inteligibilidade mais populares são os Teste de

Rhyme Diagnóstico (DRT) e Teste de Rhyme Modificado (MRT), ver [6].

Testes subjetivos são acreditados de dar a qualidade de fala "verdadeira". Contudo, o envolvimento de ouvintes humanos os torna caro e consumidores de tempo. Tais testes podem ser usados somente nos estágios finais do desenvolvimento de sistemas de comunicação por fala e não são adequados para medidas de QoS em uma base diária.

Testes objetivos usam expressões matemáticas para prognosticar qualidade de fala. Seu baixo custo, significa que eles podem ser usados para, de forma contínua, monitorar a qualidade através da rede. Duas situações de teste diferentes pode ser distinguidas:

• Intrusiva, onde ambos, o sinal original e o distorcido estão disponíveis. Isto é ilustrado na Fig. 1, onde um sinal de sinal de referência é passado adiante para um sistema sob teste, que distorce o sinal de referência. O sinal distorcido e o sinal de referência são ambos passados adiante para uma unidade de medição de intrusiva 12, que estima a medida de qualidade medida para o sinal distorcido.

• Não intrusiva (algumas vezes também denotada "término único" ou "sem referência"), onde somente o sinal distorcido está disponível. Isto é ilustrado na Fig. 2. Neste caso uma unidade de medição de não intrusiva 14 estima a medida de qualidade diretamente do sinal distorcido sem acesso ao sinal de referência.

A classe mais simples de medidas de qualidade de objetivo intrusiva são algoritmos de comparação de forma de onda, tal como relação sinal-para-ruído (SNR) e relação sinal-para-ruído em segmento (SSNR). Os algoritmos de comparação de forma de onda são simples de implementar e requerem complexidade computacional baixa, mas eles não se correlacionam bem com medidas subjetiva se tipos diferentes de distorções são comparados.

Técnicas do domínio da freqüência, tal como a medida de Itakura-Saito (IS), e a medida de distorção espectral (SD) são amplamente

usadas. Técnicas do domino da freqüência, não são sensitivas a deslocamento de tempo e são, de forma geral, mais consistentes com a percepção humana, ver [7].

Um número significativo de medidas no domínio da percepção intrusivas têm sido desenvolvido. Essas medidas incorporam o conhecimento do sistema de percepção humana. Imitação da percepção humana é usada para redução de dimensão e um estágio "cognitivo" é usado para efetuar o mapeamento para a escala de qualidade. O estágio cognitivo é treinado por meio de um ou mais bancos de dados. Essas medidas incluem a Distorção Espectral de Bark (BSD), ver [8], Qualidade de fala Perceptual (PSQM), ver [9], e Blocos de Normalização de Medidas (MNB), ver [10], [11]. Avaliação Perceptual de Qualidade de fala (PESQ), ver [12], e Avaliação Perceptual de Qualidade de Áudio (PEAQ), ver [13], são algoritmos do estado da técnica padronizados para avaliação de qualidade intrusiva de fala e áudio, respectivamente.

Medidas de qualidade de fala de objetivo intrusiva existentes podem automaticamente avaliar o desempenho do sistema de comunicação sem a necessidade de ouvintes humanos ouvintes. Contudo, medidas intrusivas requerem acesso ao sinal original, que tipicamente não está disponível na monitoração de QoS. Para tais aplicações, avaliação de qualidade não intrusiva precisa ser usada. Esses métodos freqüentemente incluem ambos, imitação de ambos, imitação de percepção humana e/ou um mapeamento para a medida de qualidade que é treinado usando os bancos de dados.

Uma tentativa anterior em direção a medida de qualidade de fala não intrusiva com base em um espectrograma do sinal percebido é apresentado em [14]. O espectrograma é dividido, e o intervalo de dinâmica e variância é calculado de uma maneira de bloco a bloco. O nível médio do intervalo de dinâmica e variância é usado para prognosticar qualidade de fala.

A avaliação de qualidade de fala não intrusiva reportada em [15] tenta prognosticar a probabilidade que a seqüência de áudio de passagem é gerada através do sistema de produção vocal humano. A seqüência de fala sob avaliação é reduzida a um conjunto de características. Os dados parametrizados são usados para estimar a qualidade percebida de forma fisiológica, por meio de regras básicas.

A medida proposta em [16] é baseada na comparação da fala emitida com um sinal artificial de referência que é, de forma apropriada, selecionado, de forma ótima, de um livro de código agregado. No Perceptual Linear Prognóstico (PLP), ver [17], coeficientes são usados como uma representação paramétrica do sinal de fala. Um modelo polinomial de quinta ordem é efetuado para suprimir detalhes dependentes de alto-falante do espectro auditivo. A distância média entre o vetor de teste desconhecido e o mais próximo centróide de referência fornece uma indicação da degradação da fala.

Algoritmos recentes com base em Modelos de probabilidade de mistura Gaussianos (GMM) de características derivadas das representações de envelope espectral motivadas pela forma de percepção podem ser encontradas em [18] e [19]. Um algoritmo novo de avaliação de qualidade motivado pela forma de percepção com base na representação de envelope temporal da fala é apresentado em [20] e [21].

O padrão de União de Telecomunicação Internacional (ITU) para avaliação de qualidade não intrusiva, ITU-T P.563, pode ser encontrado em [22]. Um total de 51 características de fala é extraído do sinal. Características principais são usadas para determinar uma classe de distorção dominante, e em cada classe de distorção, uma combinação linear das características é usada para prognosticar uma assim chamada qualidade de fala intermediária. A qualidade de fala final é estimada da qualidade intermediária e 11 características adicionais.

As medidas listadas acima para avaliação de qualidade são designadas para prognosticar os efeitos de muitos tipos de distorções, e tipicamente tem complexidade computacional alta. Tais algoritmos serão referidos como fazedores de prognósticos de qualidade de fala geral. Tem sido mostrado que prognóstico de qualidade não intrusiva é possível em complexidade muito mais baixa se é assumido que o tipo de distorção é conhecida, ver [23]. Contudo, a última classe de medidas é provável sofrer de desempenho de prognóstico pobre se as condições de trabalho esperadas não são encontradas.

SUMÁRIO

Um objeto da presente invenção é um método de avaliação de qualidade de fala não intrusiva e aparelho tendo complexidade computacional baixa.

Este objeto é alcançado de acordo com as reivindicações anexas.

A presente invenção faz prognóstico da qualidade de fala a partir das características genéricas comumente usadas em codificação de fala (referida como características por quadro), sem uma presunção do tipo de distorção. No proposto método de avaliação de qualidade de fala não intrusiva e de complexidade baixa, a estimativa de qualidade é em vez disso baseada em propriedades estatísticas globais das características por quadro.

De forma breve, a presente invenção determina parâmetros representando quadros do sinal monitorado. A coleção de vetores de característica por quadro representando informação estrutural dos quadros selecionados é extraída desses parâmetros. Um conjunto de características global é obtido da coleção de vetores de característica usando momentos estatísticos predeterminados de componentes de vetores de características selecionadas. Finalmente, uma medida de qualidade de sinal é prognosticada a partir do conjunto de características globais.

DESCRIÇÃO BREVE DOS DESENHOS

A invenção, junto com objetos e vantagens adicionais dela, pode melhor ser entendida fazendo referência a seguinte descrição tomada em conjunto com os desenhos anexos, nos quais:

Fig. 1 é um diagrama em bloco ilustrando medição de qualidade de fala intrusiva;

Fig. 2 é um diagrama em bloco ilustrando medição de qualidade de fala não intrusiva;

Fig. 3 é um diagrama em bloco ilustrando percepção humana de qualidade de fala;

Fig. 4 é um fluxograma ilustrando o método de avaliação de qualidade de sinal de acordo com a presente invenção;

Fig. 5 é um fluxograma ilustrando a modalidade preferida do método de avaliação de qualidade de sinal de acordo com a presente invenção; e

Fig. 6 é um diagrama em bloco de uma modalidade preferida do aparelho de avaliação de qualidade de sinal de acordo com a presente invenção.

DESCRIÇÃO DETALHADA

Na seguinte descrição a invenção será descrita com referência a fala. Contudo, os mesmos princípios podem ser aplicados para outros tipos de sinal, tal como sinais de áudio e sinais de vídeo.

O processo de avaliação de qualidade de fala humano pode ser dividido em duas partes: 1) conversão do sinal de fala recebido em excitação dos nervos auditivos para o cérebro, e 2) processamento cognitivo no cérebro. Isto é ilustrado na Fig. 3, onde um sinal distorcido é recebido por um bloco de processamento auditivo 16, que transforma o sinal em excitações de nervo que são passados adiante para um bloco de mapeamento cognitivo 18, que emite um sinal com uma certa qualidade percebida. O princípios principais de

transformadas de percepção são encobrimento de sinal, resolução de espectro de banda crítica, curvas de sonoridade iguais, e lei de sonoridade de intensidade, e.g., [24]. Esses princípios são bem estudados e na maioria dos algoritmos de avaliação de qualidade existentes em uma transformada de percepção é um estágio de pré-processamento. O propósito implícito principal da transformada de percepção é para efetuar uma redução de dimensão consistente da forma de percepção no sinal de fala. De forma ideal, uma transformação de percepção retém todas a informação relevante da forma de percepção, e descarta toda a informação irrelevante da forma de percepção. Na prática, aproximações e simplificações precisam ser feitas e este objetivo pode não ser encontrado. Em alguns casos, transformações de percepção podem terminal custos computacionais altos. Para evitar essas limitações em potencial, o método de avaliação da presente invenção não efetua tal uma transformada de percepção. Em vez disso a dimensionalidade é preferencialmente reduzida, de forma simultânea, com a otimização dos coeficientes da função de mapeamento. Esta meta é minimizar a perda de informação relevante. Esta abordagem é consistente com o aparecimento recente de algoritmos efetuando avaliação de qualidade sem a transformada de percepção na avaliação de qualidade de imagem [25].

Muitos dos algoritmos existentes de avaliação de qualidade existentes são baseados nos modelos específicos de distorção, i.e., nível de ruído de fundo, ruído multiplicativo, presença de tons de campainha [22], ou simulam uma distorção conhecida como características de receptor de fonte de energia de ouvido [12]. A presente invenção não incorpora um modelo explícito da distorção. A estimativa de qualidade de fala é baseada totalmente nas estatísticas de um sinal de fala processado, e a distorção é, de forma implícita, avaliada por segunda unidade de entrada impacto nessas estatísticas. Como um resultado, a presente invenção é facilmente adaptada para os sistemas de comunicação de próxima geração que provavelmente

produzir novos tipos de distorções.

Em alguns métodos, a informação dependente do alto-falante é removida [18], [16]. Contudo, é conhecido que sistemas de telefonia fornecem graduações de qualidade mais alta para algumas vozes do que para outras [26]. Por conseguinte, se o algoritmo é para ser usado em monitoramento da rede contínuo, e o material de fala balanceado para ponderação não pode ser garantido,a informação dependente de alto-falante é relevante. O método de acordo com a presente invenção incorpora a informação dependente de alto-falante, por exemplo na forma do período de ocorrência dos coeficientes de um modelo auto regressivo (AR) de décima ordem, estimado por meio de prognóstico linear.

Uma elocução usada para medições de qualidade é tipicamente um conjunto de curtas sentenças separadas por uma pausa de, por exemplo, 0,5 segundos. O comprimento total de uma elocução é tipicamente de aproximadamente 8 segundos. Contudo, em geral, uma elocução pode simplesmente ser vista como um intervalo ou bloco de sinal. O método de avaliação da presente invenção faz prognóstico de qualidade de fala de uma elocução usando um conjunto simples de características que pode ser derivado da forma de onda do sinal de fala ou, em uma modalidade preferida, estão facilmente disponíveis a partir dos codecs de fala na rede. A qualidade de fala é prognosticada em complexidade computacional baixa, que torna o método útil para aplicações práticas.

O núcleo do método de avaliação de qualidade de sinal de acordo com a presente invenção é um vetor de características $\Phi(n)$ por quadro de múltiplas dimensões (preferencialmente de Expressão 11 dimensões para fala; outros números também são possíveis e o número de dimensões também depende do tipo de sinal, fala, áudio, vídeo, etc), os componentes dos quais estão definidas no APÊNDICE I. A qualidade de fala não é prognosticada diretamente a partir do vetor por quadro, mas de suas propriedades estatísticas

globais, descrita como média, variância, inclinação, e kurtosis das características por quadro sobre muitos quadros, por exemplo sobre uma elocução. As propriedades estatísticas das características por quadro (referidas como conjunto de características global Ψ) formam a entrada para o mapeamento de GMM (Modelo de Probabilidade de Mistura Gaussiano), que estima o nível de qualidade de fala sobre uma escala de MOS, como descrita em detalhe no APÊNDICE III.

Fig. 4 é um fluxograma ilustrando o método de avaliação de qualidade de sinal de acordo com a presente invenção. No estágio S1 um sinal de fala é codificado em uma seqüência de bit de quadros incluindo parâmetros de fala. Estágio S2 extrai uma vetor de característica Φ(n) (por quadro) dos parâmetros de fala para cada quadro de interesse. No estágio S3 as propriedades estatísticas desses vetores de característica são usadas para formar um conjunto de característica Ψ global (por elocução). Finalmente, no estágio S4 a qualidade de fala é prognosticada a partir do conjunto de características global usando mapeamento de GMM.

A base do aparelho e método de avaliação de qualidade de sinal de acordo com a presente invenção é a extração de um vetor de característica. O conjunto de características usado auxilia a capturar a informação estrutural a partir de um sinal de fala. Isto é motivado pelo fato que o sinal de fala natura é altamente estruturado, e é provável que o julgamento de qualidade humano se baseie em padrões extraídos da informação descrevendo essa estrutura. APÊNDICE I define um conjunto of 11 características adequadas, que são coletadas em um vetor de característica por quadro:

$$\Phi(n) = (\Phi_1(n), \Phi_2(n), \ldots \Phi_{11}(n)) \quad (1)$$

onde n denota o número do quadro.

De acordo com a presente invenção é assumido que a qualidade de fala pode ser estimada a partir das propriedades estatísticas

dessas características por quadro. Suas distribuições de probabilidade são descritas com a média, variância, inclinação, e kurtosis. Esses momentos estatísticos são calculados, de forma independente para cada característica por quadro, e isto fornece um conjunto de características que, de forma global descrevem uma elocução de fala (características globais):

$$\mu_{\Phi_i} = \frac{1}{|\tilde{\Omega}|}\sum_{n\in\tilde{\Omega}} \Phi_i(n) \tag{2}$$

$$\sigma_{\Phi_i} = \frac{1}{|\tilde{\Omega}|}\sum_{n\in\tilde{\Omega}} \left(\Phi_i(n) - \mu_{\Phi_i}\right)^2 \tag{3}$$

$$s_{\Phi_i} = \frac{1}{|\tilde{\Omega}|}\frac{\sum_{n\in\tilde{\Omega}} \left(\Phi_i(n) - \mu_{\Phi_i}\right)^3}{\sigma_{\Phi_i}^{3/2}} \tag{4}$$

$$k_{\Phi_i} = \frac{1}{|\tilde{\Omega}|}\frac{\sum_{n\in\tilde{\Omega}} \left(\Phi_i(n) - \mu_{\Phi_i}\right)^4}{\sigma_{\Phi_i}^{2}} \tag{5}$$

Aqui $\tilde{\Omega}$ denota o conjunto de quadros, de cardinalidade (tamanho) $|\tilde{\Omega}|$, usado para calcular estatísticas para cada uma das características por quadro Φ_i (n). As características globais são agrupadas em um conjunto de características globais:

$$\Psi = \left\{\mu_{\Phi_i}, \sigma_{\Phi_i}, s_{\Phi_i}, k_{\Phi_i}\right\}_{i=1}^{11}$$

Preferencialmente a complexidade desses cálculos é reduzida. APÊNDICE II descreve um procedimento de redução de dimensionalidade em dois estágios que:

- Extrai o "melhor" sub-conjunto $\tilde{\Omega}$ de quadros fora do conjunto de todos os quadros na elocução.
- Transforma conjunto de características globais Ψ em um conjunto de características globais Ψ de dimensionalidade mais baixa.

Em uma modalidade preferida da presente invenção a características por quadro do enésimo quadro são calculados diretamente da variância E^e da excitação do modelo de AR, o período de ocorrência T e o vetor de dez dimensões dos coeficientes f de freqüência espectral em linha (LSF), determinada sobre 20 ms de quadros de fala. Já que E^e, T e f são prontamente acessíveis na rede no caso de codificadores de Prognóstico Linear Excitado por Código (CELP) [27], esta modalidade de uma invenção tem a vantagem adicional de extrair o vetor por quadro diretamente dos parâmetros de rede na seqüência de bit, que é mais implementações do que extraí-lo da forma de onda do sinal. Será demonstrado que as características por quadro $\Phi(n)$ podem ser calculadas de $\{E^e_n, T_n, f_n\}$ e $\{E^e_{n-1}, T_{n-1}, f_n\}$. Então será mostrado como as propriedades estatísticas globais são calculadas, de forma recursiva, sem armazenar as características por quadro para a elocução inteira em uma área de armazenamento temporário. O período de ocorrência T_n é calculado de acordo com [40], e os coeficientes de AR são extraídos do sinal de fala a cada 20 ms sem sobrepor.

Para manter a complexidade do método baixa, as características por quadro: planicidade espectral, dinâmica espectral, e centróide espectral são aproximadas. As aproximações são baseadas inteiramente na seqüência de bit codificada de fala, e por meio disso a reconstrução do sinal é evitada.

Em uma modalidade preferida da presente invenção a planicidade espectral é aproximada conforme a proporção da variância de erro de prognóstico de décima ordem e a variância do sinal:

$$\Phi_1(n) = \frac{E_n^e}{E_n^s} \qquad (7)$$

Dada a variância da excitação do modelo de AR, sua definição

$$e_k = s_k - \sum_{i=1}^{10} a_i s_{k-i} \qquad (8)$$

e coeficientes de AR a_i, o sinal de variância é calculado sem reconstruir a forma de onda s_k usando a recursão reversa de Levinson-Durbin (algoritmo de estágio abaixo).

As dinâmicas espectrais são preferencialmente aproximadas por uma descrição paramétrica do envelope do espectro, por exemplo como uma distância ponderada Euclideana no espaço de LSF:

$$\Phi_2(n) = (\mathbf{f}_n - \mathbf{f}_{n-1})^T \mathbf{W}_n (\mathbf{f}_n - \mathbf{f}_{n-1}) \qquad (9)$$

onde a ponderação média de harmônico inversa [41] é definida pelos componentes do vetor de LSF:

$$W^{(ii)} = \left(f_n^{(i)} - f_n^{(i-1)}\right)^{-1} + \left(f_n^{(i+)} - f_n^{(i)}\right)^{-1} \qquad (10)$$
$$W^{(iv)} = 0$$

Em uma modalidade preferida da presente invenção as médias são também usadas para obter um centróide espectral aproximado:

$$\Phi_3(n) = \frac{\sum_{i=1}^{10} i W_n^{(ii)}}{\sum_{i=1}^{10} W_n^{(ii)}} \qquad (11)$$

As características globais selecionadas são calculadas, de forma recursiva, i.e., as características por quadros não são armazenada em uma área de armazenamento temporário. Até o fim da elocução, a média é, de forma recursiva atualizada de acordo com:

$$\mu_\Phi(n) = \frac{n-1}{n}\mu_\Phi(n-1) + \frac{1}{n}\Phi(n) \qquad (12)$$

para obter a μ_Φ. aqui n é o índice sobre o conjunto de quadro aceito $\tilde{\Omega}$, como discutido no APÊNDICE II. Em um modo similar, Φ^2, Φ^3 e Φ^4 são propagados para obter os momentos centrais $\mu_\Phi{}^2$, $\mu_\Phi{}^3$ e $\mu_\Phi{}^4$. Essas quantidades são usadas para obter as características globais remanescentes, a saber variância, inclinação, e kurtosis como:

$$\sigma_\Phi = \mu_{\Phi^2} - (\mu_\Phi)^2$$
$$s_\Phi = \frac{\mu_{\Phi^3} - 3\mu_\Phi\mu_{\Phi^2} + 2(\mu_\Phi)^3}{\sigma_\Phi^{3/2}} \qquad (13)$$
$$k_\Phi = \frac{\mu_{\Phi^4} - 4\mu_\Phi\mu_{\Phi^3} + 6(\mu_\Phi)^2\mu_{\Phi^2} - 3(\mu_\Phi)^4}{\sigma_\Phi^2}$$

A modalidade preferida do método de acordo com a presente invenção é ilustrada na Fig. 5. Ela inclui os seguintes estágios:

S5. Para o enésimo quadro de fala determina $\{E^e_n, T_n, f_n\}$ e $\{E^e_{n-1}, T_n, f_{n-1}\}$ da forma de onda ou extraí da seqüência de bit.

S6. Determinar vetor de característica por quadro $\Phi(n)$, com base em $\{E^e_n, T_n, f_n\}$ e os correspondentes parâmetros $\{E^e_{n-1}, T_n, f_{n-1}\}$ do quadro anterior, que são armazenadas em uma área de armazenamento temporário.

Estágios S5 e S6 são efetuados para todos os quadros da elocução.

S7. A partir de um sub-conjunto selecionado $\tilde{\Omega}$ de quadros, de forma recursiva, determinar os momentos centrais $\{ \mu_\Phi, \mu_\Phi^2, \mu_\Phi^3$ e $\mu_\Phi^4\}$. Quadro Seleção de quadro (APÊNDICE II) é controlada pelo limite ou janela de várias dimensões Θ.

S8. Ao final da elocução calcular o conjunto de características global $\tilde{\Psi} = \{ \mu_{\Phi i}, \sigma_{\Phi i}, s_{\Phi i}, k_{\Phi i} \}$ selecionado (equação (23) of APÊNDICE II) como média, variância, inclinação, e kurtosis das características por quadro.

S9. Prognosticar a qualidade de fala da elocução como a função do conjunto de características global $Q = Q(\Psi)$ através de mapeamento de GMM, conforme descrito no APÊNDICE III.

Fig. 6 é um diagrama em bloco de uma modalidade preferida do aparelho de avaliação de qualidade do sinal de acordo com a presente invenção. Um sinal codificado de fala (seqüência de bit) é recebido através de um calculador de vetor de características 60, que determina o vetor de características $\Phi(n)$ do quadro corrente n dos parâmetros de fala $\{E^e_n, T_n, f_n\}$.

O vetor de características é passado adiante para um seletor de quadro 62, que determina se ele está dentro da janela de várias dimensões definida pelo limite Θ, que é armazenada no depósito 64 e tem sido determinada por tentativa como descrito no APÊNDICE II e IV. Os componentes do vetor de características Φ_1 (n), Φ_2 (n),...Φ_{11} (n) dos quadros selecionados são passados adiante para os respectivos calculadores 66, 68, 70, 72, que, de forma recursiva, calculam momentos centrais de cada componente. Na Fig. 6 somente os calculadores para Φ_1 tem sido ilustrados de forma explícita, os correspondentes calculadores para os componentes remanescentes tem sido indicados por pontos. Os momentos centrais são passados adiante para os calculadores de características globais 74, que determinam as características globais de cada componente de vetor de características de acordo com a equação (13). O conjunto de características globais resultante é passado adiante para um previsor de qualidade, que determina uma estimativa de qualidade Q como descrito no APÊNDICE III.

Efetivamente, em uma implementação prática, todos os calculadores de momentos centrais 66, 68, 79, 72 pode não ser exigidos para cada componente de vetor de características Φ_i, já que a redução da dimensionalidade do conjunto de características globais Ψ pode reduzir o número de características globais requeridas, como ilustrado pela equação (23) no APÊNDICE II. Neste caso, calculador 72 para componente de vetor de características Φ_i pode ser omitido, já que $k_{\Phi i}$, tem sido descartado na redução de dimensionalidade do conjunto de características global Ψ. Os momentos centrais requeridos efetivamente dependem dos resultados da redução de dimensionalidade do conjunto de características global Ψ, que por sua vez também depende do tipo de sinal tipo (fala, áudio, vídeo, etc).

A funcionalidade do aparelho de avaliação da presente invenção é tipicamente implementada por um micro processador ou por uma combinação micro/processador de sinal e software correspondente.

Embora o prognóstico de qualidade efetuado pela presente invenção seja baseado no mapeamento de modelo de probabilidade de Mistura Gaussiana, outras alternativas factíveis são os modelos de redes neural e de Markov escondido.

O desempenho do método de avaliação de qualidade de acordo com a presente invenção tem sido avaliado, de forma experimental. Os resultados são dados no APÊNDICE V.

Um aspecto não coberto, de forma explícita, na descrição acima são os parâmetros afetando a qualidade de fala, embora não diretamente acoplados ao sinal de fala original. Tais parâmetros são, e.g. ruído de fundo para a entrada de fala ou tipo de alto-falante (e.g. gênero) ou entrada sinais não de fala como música, ou o estilo de música (e.g. pop, jazz, clássico,...), ou parâmetros relacionados ao sistema de transmissão, tais como:

- O uso de um sistema de VAD /DTX para transmissão eficiente de fala inativa
- O uso (existência) de um supressor de ruído antes ou no decorres da codificação de fala
- O método de codificação de fala e sua configuração (codec selecionado e seu modo ou taxa de bit)
- Indicadores de quadro ruim indicando que um quadro do codec é, não usável, de forma parcial, ou completamente, devido a erros de transmissão
- Um parâmetro de probabilidade para a ocorrência de erros de transmissão na seqüência de bit de fala recebida, que pode ser derivado de vários estágios de processamento no receptor.
- Um possível codec em tandem envolvendo uma multidão de decodificações e re-codificações do sinal de fala
- Modificação da escala de tempo possível da fala em conjunto com o uso de uma área de armazenamento temporário de variação adaptativa.

Esses parâmetros têm uma imediata influência na qualidade de fala resultante após a decodificação. A aplicação direta da presente invenção vai ignorar esses parâmetros, que conduzem para a vantagem de um método de avaliação de qualidade universal com complexidade baixa.

Contudo, pelo menos, algum desses parâmetros são conhecidos ou podem ser conhecidos a priori ou podem ser deduzidos usando detectores correspondentes ou podem ser obtidos por meio de sinalização através do sistema de transmissão de fala. Como um exemplo, condições de música ou de ruído de fundo ou a existência da supressão de ruído podem ser detectadas usando métodos de detecção do estado da técnica. Meios de sinalização são adequados para identificar os outros mencionados parâmetros. Uma modalidade específica da invenção fazendo uso dos parâmetros a prioridade é para usar vários exemplos do método de avaliação de qualidade de acordo com a presente invenção, que são treinados para diferentes conjuntos desses parâmetros. De acordo com esta modalidade o exemplo do método de avaliação que é mais adequado presentemente para um dado conjunto de parâmetros a priori é primeiro identificado e selecionado. Em um segundo estágio o exemplo selecionado é executado rendendo a estimativa de qualidade de fala desejada.

Uma modalidade adicional é executar um exemplo único de um método de avaliação de qualidade de acordo com a presente invenção, seguido de um estágio de processamento adicional levando em conta os parâmetros a priori. De forma específica, o segundo estágio pode efetuar um mapeamento do valor de saída do método de avaliação do primeiro estágio e dos vários parâmetros a priori para a saída final da estimativa da qualidade de fala. O mapeamento deste segundo estágio pode ser feito de acordo com técnicas conhecidas tal um método de adequação de dados de quadrados mínimos linear ou não linear ou mapeamentos de GMM. Mesmo uma possibilidade adicional seja combinar o estágio de mapeamento final de

GMM de um método de avaliação de qualidade com o mapeamento de segundo estágio descrito, que essencialmente estende o vetor de características globais (por elocução) através do conjunto dos parâmetros a priori.

Ainda uma modalidade adicional para tornar o método mais aplicável para não fala, e música em particular, é permitir adaptações das características ' por quadro" locais usadas. Música em geral não é bem codificada com codecs de fala, já que música não vai de encontro ao modelo de produção de fala existentes dos codecs de fala. Mais propriamente, música é preferencialmente codificada com base nos modelos de percepção (da audição), não assumindo qualquer modelo particular da produção de sinal fonte. Considerando este fato, uma adaptação das características "por quadro" locais significa, preferencialmente, usar parâmetros derivados de tal um modelo de percepção, pelo menos, em adição aos parâmetros apresentados. Particularmente, este é o caso se o codec usado é um codec de áudio mais propriamente do que codec de fala.

Um outro aspecto é que a descrição acima descreve a invenção em uma forma de regressão, que efetua um mapeamento contínuo. Contudo, o método é também aplicável para um mapeamento discreto para pré-definir a escala de qualidade discreta (intervalos pré-definidos) por meio do uso de um classificador. Então, o termo "mapeamento" deves em resposta interpretado em um sentido geral que também cobre o caso discreto de usar um classificador. Um exemplo simples com um classificar é um sistema com base no método de avaliação de qualidade descrito que não faz prognóstico de qualidade em uma escala contínua, mas tem um resultado binário, por exemplo: 0) qualidade está abaixo de um limite e 1) qualidade está acima de um limite. Este exemplo corresponde a um sistema que tem a habilidade de detectar se uma particular distorção ou nível de qualidade está ou não presente.

As várias modalidades da presente invenção conduzem a uma ou várias das seguintes vantagens:

• Qualidade de fala pode ser prognosticada a partir dos parâmetros da seqüência de bit (em um caso de codecs CELP), sem reconstrução da forma de onda. Isto, junto com o fato que transformada para o domínio da percepção não usado, conduz a requisitos baixos de computação e de memória (a complexidade de umas poucas centenas de vezes mais baixa do que o padrão ITU existente).

• A qualidade de fala é prognosticada a partir das propriedades estatísticas das características: planicidade espectral, centróide espectral, dinâmica espectral, período de ocorrência, sinal variância de sinal, variância do sinal de excitação sinal, e outras derivadas no tempo. As propriedades estatísticas dessas características são descritas por meio sua média, variância, inclinação, e kurtosis. Este tipo de característica e sua descrição não requerem que o sinal de fala seja armazenado. Em uma área de armazenamento temporário são armazenados somente uns poucos (por exemplo 12) parâmetros escalares do quadro anterior.

• Um novo método pode ser usado para derivar as características por quadro (planicidade espectral, dinâmica espectral, etc.) diretamente da seqüência de bit, sem reconstruir a forma de onda (reconstrução do sinal não é não complexa por si só, a complexidade ocorre quando as características são extraídas do sinal reconstruído).

• A qualidade de fala pode ser prognosticada a partir de somente um sub-conjunto de quadros. Um novo método é usado para extrair os quadros que contem informação útil (nos método existente de avaliação de qualidade, rejeição de quadro é baseado em um simples limite de energia ou detector de atividade de fala). O método proposto generaliza esta abordagem. Diferentes sub conjuntos de quadros podem ser usados para estimar as propriedades estatísticas de características diferentes. Rejeição de quadro não

é somente uma função de energia, mas de todas as características por quadro. O método de seleção de quadro pode ser otimizado em conjunto com a função de regressão (classificador).

• O método proposto, de forma significativa, executa a ITU-T P.563 nas simulações formadas, com relação ao coeficiente de correlação e ao erro de raiz quadra média.

Será entendido por aqueles com qualificação na arte que várias modificações e mudanças podem ser feitas para a presente invenção sem fugir do escopo da dela, que é definido pelas reivindicações anexas.

APÊNDICE I

SELEÇÃO DE CARACTERÍSTICA

Este apêndice vai definir um conjunto adequado de características a ser incluído em um vetor de características por quadro Φ(n). este conjunto é especificamente adequado para sinal de fala.

Uma primeira característica por quadro de interesse é a medida representando o conteúdo de informação no sinal, tal como a medida de planicidade espectral descrita em [28]. Esta é relacionada com a força da estrutura ressonante no espectro de potência e é definida como:

$$\Phi_1(n) = \frac{\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(P_n(\omega)\right)d\omega\right)}{\frac{1}{2\pi}\int_{-\pi}^{\pi}P_n(\omega)d\omega} \qquad (14)$$

(ω)

onde o envelope de AR (Auto-Regressivo) $P^{(\omega)}$ é definido como a resposta de freqüência do modelo de AR modelo com coeficientes a_k, i.e.

$$P_n(\omega) = \frac{1}{\left|1+\sum_{k=1}^{p}a_k^{(n)}e^{-j\omega k}\right|^2} \qquad (15)$$

O índice de quadro é denotado por n, e p é a ordem da análise

de prognóstico linear, tipicamente configurado to 10 para sinais amostrados em 8 kHz.

Uma segunda característica por quadro é a medida representando a estabilidade do sinal, tal como a dinâmica espectral, definida como:

$$\Phi_2(n)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\log\left(P_n(\omega)\right)-\log\left(P_{n-1}(\omega)\right)\right)^2 d\omega \qquad (16)$$

A característica da dinâmica espectral tem sido estudadas e com sucesso usada em codificação de fala [29], [30] e aprimoramento de fala [31].

Uma terceira característica de interesse é a medida representando a distribuição de energia do sinal sobre freqüências, tal como o centróide espectral [32], que determina a área de freqüência em torno da qual a maioria da energia do sinal se concentra. Isto é definido como:

$$\Phi_3(n)=\frac{\frac{1}{2\pi}\int_{-\pi}^{\pi}\omega\log\left(P_n(\omega)\right)d\omega}{\frac{1}{2\pi}\int_{-\pi}^{\pi}P_n(\omega)d\omega} \qquad (17)$$

e isto é também freqüentemente usado como uma aproximação de uma medida de "luminosidade" de percepção.

Três características por quadro adicionais são a variância da excitação do modelo de AR modelo $E^e_{n,}$, a variância do sinal de fala variância E^s_n, e o período de ocorrência $T_{n,}$. Elas serão denotadas como Φ_4 (n), Φ_5 (n), e Φ_6 (n), respectivamente.

Φ_4 (n) = E^e_n, a variância da excitação do modelo de AR

Φ_5 (n) = E^e_n, a variância do sinal de fala (18)

Φ_6 (n) = T_n, o período de ocorrência

As características por quadro apresentadas acima, e suas primeiras derivadas no tempo (excetos as derivadas das dinâmicas espectrais) são agrupadas em um vetor de características por quadro de 11 dimensões

$\Phi(n)$, os componentes dos quais são resumidos na Tabela I abaixo.

Tabela 1

Elementos do vetor de características por quadro

Descrição	Característica	Derivada no tempo da característica
Planicidade espectral	Φ_1	Φ_7
Dinâmica espectral	Φ_2	-
Centróide espectral	Φ_3	Φ_8
Variância de excitação	Φ_4	Φ_9
Variância de fala	Φ_5	Φ_{10}
Período de passo	Φ_6	Φ_{11}

APÊNDICE II

REDUÇÃO DE DIMENSÃOALIDADE

Redução de dimensionalidade pode ser obtida através da seleção de quadro, global seleção de características globais ou uma combinação de ambos os procedimentos de seleção. Um propósito da redução de dimensionalidade é melhorar a precisão de prognóstico do sistema de avaliação de qualidade removendo dados irrelevantes e redundantes. Um outro propósito é reduzir a complexidade computacional. A redução de dimensionalidade apresentada neste apêndice é baseada em um procedimento de treinamento que será descrito em detalhes no APÊNDICE IV.

Uma comumente abordagem usada na literatura de avaliação de qualidade, é remover regiões de não fala com base em um detector de atividade de fala ou um limite de energia [33]. A presente invenção sugere uma generalização deste conceito considerando limites de atividade em todas as dimensões de característica por quadro.

O esquema, apresentado no método de seleção de quadro abaixo permite quadros ativos de fala a serem excluídos se eles não transportam informação que melhora a precisão do prognóstico de qualidade de fala. O conceito do algoritmo de seleção de quadro é aceitar somente quadros onde o vetor de características por quadro $\Phi(n)$ se encontra dentre ou

na superfície do "hyperbox" de 11 dimensões ou janela de várias dimensões definida por um vetor limite vetor Θ. Em pseudo código o método pode ser descrito por:

MÉTODO DE SELEÇAO DE QUADRO

$\tilde{\Omega} = \{Ø\}$ Configurar subconjunto $\tilde{\Omega}$ para conjunto vazio

para n € Ω para cada quadro no conjunto de quadro original Ω

se Φ_1 (n) € [Θ_1^L, Θ_1^U] & se vetor de características está dentro da "janela"

Φ_2 (n) e [Θ_2^L, Θ_2^U] &

.

.

.

Φ_{11} (n) e [Θ_{11}^L, Θ_{11}^U]

Então $\tilde{\Omega} = \tilde{\Omega} + \{n\}$ Adiciona quadro n ao subconjunto $\tilde{\Omega}$

O conjunto ótimo de quadros é determinado pelo limite ou janela de várias dimensões $\Theta = \{ \Theta_1^L, \Theta_1^U \}_{i=11}^{11}$ i.e. $\tilde{\Omega}$ depende de Θ, ou $\tilde{\Omega} = \tilde{\Omega}(\Theta)$. Nos procuramos pelo limite Θ que minimiza o critério ε:

$$\Theta = \arg\min_{\Theta^*} \varepsilon\left(\tilde{\Omega}(\Theta^*)\right) \qquad (19)$$

O critério ε é calculado como o desempenho de erro da média da raiz quadrada (RMSE) do método de avaliação de qualidade de acordo com a presente invenção, i.e.:

$$\varepsilon = \sqrt{\frac{\sum_{i=1}^{N}\left(Q_i - \hat{Q}_i\right)^2}{N}} \qquad (20)$$

onde $\hat{Q}$ é a qualidade prognosticada, e é Q a qualidade subjetiva. Aqui N é o número de elocuções rotuladas de MOS usado na avaliação, ver APÊNDICE IV. A otimização do limite Θ é baseado no conjunto inteiro de características globais Ψ. A otimização de ε em (19), com o algoritmo de seleção de quadro descrito acima, resultas no seguinte critério para a aceitação do enésimo quadro:

$$\Phi_5(n) > \Theta_5^L \;\&\; \Phi_1(n) < \Theta_1^U \;\&\; \Phi_2(n) < \Theta_2^U \qquad (21)$$

com os valores limite $\Theta_5^L = 3{,}10$, $\Theta_1^U = 0.67$, e $\Theta_2^U = 4{,}21$.

De (21) é visto que somente três características por quadro têm impacto significativo na seleção de quadro, a saber variância de fala Φ_5, planicidade espectral Φ_1, e dinâmica espectral Φ_2. A primeira e segunda desigualdade em (21) somente aceitam quadros com alta energia e uma estrutura de formação clara. Isto sugere que um algoritmo de avaliação de qualidade da presente invenção extrai informação sobre a qualidade de fala predominantemente de regiões de fala manifestadas. A terceira desigualdade somente seleciona regiões de fala estacionárias. O último resultado é provavelmente devido a distorção sendo mais facilmente percebida nas regiões de repouso do sinal de fala.

Como pode ser visto do critério (21), o limite ou janela de várias dimensões pode ter restrições efetivas somente em poucas dimensões. Nas outras dimensões a janela pode ser considerada como uma janela infinita. Ainda mais, mesmo em dimensões restritas, a janela pode ter uma fronteira ou limite somente em uma direção. Em geral a janela de várias dimensões é mais restritiva na seleção de quadro do que o detector de atividade de fala pura, que conduz a rejeição de mais quadros sem informação relevante para a avaliação da qualidade. Este por sua vez conduz a médias de qualidade mais confiáveis.

Uma alternativa viável para a janela retangular para cada dimensão é uma janela suave, por exemplo uma janela Gaussiana, com cada função Gaussiana tendo sua média e variância individual. Cada componente de vetor vai então corresponder a um valor de função Gaussiana. Um quadro seria aceito se o produto desses valores de função excede um certo limite.

O critério (21) reduz, de forma significativa, o número de quadros processados através de um algoritmo de avaliação de qualidade. O

número de quadros selecionados varia com alto-falantes e sentenças, e tipicamente $\tilde{\Omega}$ contém entre 20% e 50% do conjunto de quadro total Ω.

Uma vez que o subconjunto ótimo de quadros $\tilde{\Omega}$ foi encontrado, uma procura pelo subconjunto ótimo de características globais Ψ pode ser efetuada. este estágio de otimização é definido como a seguir: dado o conjunto original de características globais Ψ de cardinalidade | Ψ |, e o conjunto ótimo de quadros $\tilde{\Omega}$, selecionar um subconjunto de características globais $\tilde{\Psi}$ C Ψ de cardinalidade $|\tilde{\Psi}| < |\Psi|$ que é otimizada para o desempenho de um algoritmo de avaliação de qualidade:

$$\tilde{\Psi} = \arg\min_{\tilde{\Psi}^* \in \Psi} \varepsilon\left(\tilde{\Psi}^*\right) \qquad (22)$$

Um procura total é somente um procedimento de redução de dimensionalidade que garante que uma ótima global seja encontrada. Contudo, é raramente aplicado devido a seus requisitos computacionais. A bem conhecida Seleção para frente Seqüencial e Seleção para trás Seqüencial, e.g., [34] são estágios [ótimo somente, já que a melhor (pior) característica global é adicionada (descartada), mas a decisão não pode ser corrigida em um estágio mais tarde. O algoritmo mais avançado (L,R) [35] consiste em aplicar Seleção para frente Seqüencial L times, seguido de R estágios de Seleção para trás Seqüencial. Os métodos de Busca Flutuante [36] são extensões dos (L,R) métodos de procura, onde o número estágios para à frente e para trás não são pré-definidos, mas obtidos de forma dinâmica. Em nossas experiência nos temos usado o procedimento de Seleção para trás Flutuante Seqüencial, que consiste em aplicar após cada estágio para trás um número de estágios para à frente enquanto os subconjuntos resultantes forem melhores do que aqueles anteriormente avaliados, como ilustrado pelo seguinte método:

PROCEDIMENTO DE SELEÇÃO PARA TRÁS FLUTUANTE SEQÜENCIAL

$\tilde{\Psi} = \Psi$	Configura o conjunto inteiro de características globais
enquanto erro não aumenta	(por mais do que um primeiro limite)
$\Psi_{i-} = \arg\min_{\Psi_i \in \tilde{\Psi}} \varepsilon\left(\tilde{\Psi} - \{\Psi_i\}\right)$	Achar a característica global menos significante
$\tilde{\Psi} = \tilde{\Psi} - \{\Psi_{i-}\}$	Excluir a característica
enquanto erro diminui	(por mais do que um segundo limite)
$\Psi_{i+} = \arg\min_{\Psi_i \in \tilde{\Psi}} \varepsilon\left(\tilde{\Psi} + \{\Psi_i\}\right)$	Achar a característica global mais significante
$\tilde{\Psi} = \tilde{\Psi} + \{\Psi_{i+}\}$	Incluir a característica

Após a otimização de ε. em (22), a dimensionalidade do conjunto de características globais é reduzida de 44 to 14, i.e. | $\tilde{\Psi}$ | = 14, e esses elementos são:

$$\tilde{\Psi} = \left\{s_{\Phi_1}, \sigma_{\Phi_2}, \mu_{\Phi_4}, \mu_{\Phi_5}, \sigma_{\Phi_5}, s_{\Phi_5}, \mu_{\Phi_6}, s_{\Phi_7}, \mu_{\Phi_8}, \mu_{\Phi_9}, \sigma_{\Phi_9}, s_{\Phi_9}, \mu_{\Phi_{10}}, \mu_{\Phi_{11}}\right\} \quad (23)$$

É notado todas as características por quadro estão presentes (através de suas representações estatística das características globais) no conjunto $\tilde{\Psi}$, mas a variância de sinal de fala Φ_5, e as derivadas da variância do sinal de excitação, Φ_9, são mais freqüentes. Uma outra observação é que as características globais de fala com base somente nos primeiros três momentos estão presentes, e que as características globais com base na kurtosis parecem ser menos importante.

APÊNDICE III

ESTIMAITVA DE QUALIDADE

Deixe Q denotar a qualidade subjetiva de uma elocução como obtida do banco de dados de treinamento rotulado de MOS. Construir um elaborador de estimativa de objetivo $\hat{Q}$ da qualidade subjetiva como uma função de um conjunto de características globais, i.e. $\hat{Q} = \hat{Q}(\tilde{\Psi})$, e procura para a função mais perto da qualidade subjetiva com relação ao critério:

$$\hat{Q}(\tilde{\Psi}) = \arg\min_{Q^*(\tilde{\Psi})} E\left\{\left(Q - Q^*(\tilde{\Psi})\right)^2\right\} \quad (24)$$

onde E{ } é o operador de expectativa. O critério definido acima é a medida probabilística correspondendo a (22) no APÊNDICE II. É bem conhecido, e.g., [37], que equação (24) é minimizada pela expectativa condicional:

$$\hat{Q}\left(\tilde{\Psi}\right) = E\left\{Q\middle|\tilde{\Psi}\right\} \qquad (25)$$

e o problema reduz para a estimativa da probabilidade condicional. Para facilitar esta estimativa, a densidade conjunta das variáveis de características globais com as graduações de MOS subjetivas podem ser modeladas como um GMM (Modelo de Probabilidade de Mistura Gaussiano):

$$f\left(\varphi\middle|\lambda\right) = \sum_{m=1}^{M} \omega^{(m)} N\left(\varphi\middle|\mu^{(m)}, \Sigma^{(m)}\right) \qquad (26)$$

onde φ = [Q, $\tilde{\Psi}$], m é o índice de componente de mistura, $\omega^{(m)}$ são os pesos da mistura, e N ($\varphi|\mu^{(m)}$, $\sum^{(m)}$) são densidades variadas Gaussianas, com $\mu^{(m)}$, $\sum^{(m)}$ sendo as matrizes de média e a de covariância das densidades Gaussianas, respectivamente. O GMM é completamente especificado por um conjunto de vetores de média, matrizes de covariância e pesos de mistura:

$$\lambda = \left\{\omega^{(m)}, \mu^{(m)}, \Sigma^{(m)}\right\}_{m=1}^{M} \qquad (27)$$

e esses coeficientes são estimados fora de linha a partir de um conjunto de treinamento grande usando um algoritmo de maximização de expectativa (EM) [38]. Detalhes sobre os dados usados para exercício são apresentados no APÊNDICE IV. Experimentos tem mostrado que é suficiente usar 12 matrizes de covariância completa (14 x 14), i.e., para dimensionalidade K = 14 e M = 12 Gaussianas, isto corresponde a M(1 + K + K(K + 1) / 2) = 1440 parâmetros de treinamento.

Usando o modelo de mistura Gaussiano de junção, a expectativa condicional (25) pode ser expressa como uma soma ponderada de

expectativas condicionais de componente inteligente, que é uma bem conhecida propriedade do caso Gaussiano [39]. Então, o elaborador de estimativa de qualidade ótima (25) pode ser expressa como:

$$\hat{Q}(\tilde{\Psi}) = E\{Q|\tilde{\Psi}\} = \sum_{m=1}^{M} u^{(m)}(\tilde{\Psi})\mu_{Q|\tilde{\Psi}}^{(m)} \quad (28)$$

onde

$$u^{(m)}(\tilde{\Psi}) = \frac{\omega^{(m)} N\left(\tilde{\Psi}\middle|\mu_{\tilde{\Psi}}^{(m)}, \Sigma_{\tilde{\Psi}\tilde{\Psi}}^{(m)}\right)}{\sum_{k=1}^{M} \omega^{(k)} N\left(\tilde{\Psi}\middle|\mu_{\tilde{\Psi}}^{(k)}, \Sigma_{\tilde{\Psi}\tilde{\Psi}}^{(k)}\right)} \quad (29)$$

e

$$\mu_{Q|\tilde{\Psi}}^{(m)} = \mu_{Q}^{(m)} + \Sigma_{\tilde{\Psi}Q}^{(m)}\left(\Sigma_{\tilde{\Psi}\tilde{\Psi}}^{(m)}\right)^{-1}\left(\tilde{\Psi} - \mu_{\tilde{\Psi}}^{(m)}\right) \quad (30)$$

com $\mu_{\Phi}{}^{(m)}$, $\mu_{Q}{}^{(m)}$, $\sum_{\Psi\,\Psi}{}^{(m)}$, $\sum_{\Psi Q}{}^{(m)}$ sendo as matrizes de média, covariância e covariância cruzada de Ψ e Q no m-ésimo componente de mistura.

APÊNDICE IV

TREINAMENTO

Para o procedimento de avaliação e treinamento nos usamos 11 banco de dados rotulados de MOS fornecidos por Ericsson AB e 7 bancos de dados rotulados de forma similar do ITU-T P.Supp 23 [43]. Dados com graduações de DMOS foram excluídas a partir dos nossos experimentos, e.g., do ITU-T P.Supp 23 nós excluímos o Experimento 2. O material de fala nesses bancos de dados contém elocuções nos seguintes idiomas: Inglês, Francês, Japonês, Italiano e Sueco. Os bancos de dados contém uma grande variedade de distorções, tal como: codificação diferentes, em tandem, e condições de unidade de referência de ruído modulada (MNRU) [44], assim como perda de pacote, ruído de fundo, efeitos de supressão de ruído, efeitos de comutação, níveis de entrada diferentes, etc. O tamanho total da união de bancos de dados é 7646 elocuções com comprimento médio de 8s.

Divide-se os bancos de dados disponíveis em duas partes,

conjunto de teste e conjunto de conjunto de treinamento. O conjunto de teste é baseado em 7 bancos de dados do ITUT P.Supp 23 (1328 elocuções), e o conjunto de treinamento é baseado em 11 bancos de dados (6318 elocuções) da Ericsson. O conjunto de teste não está disponível durante o treinamento, mas usado somente para avaliação. O treinamento usada para a redução de esquema de dimensionalidade e experimentos de avaliação de desempenho é totalmente baseada no conjunto de treinamento. Para melhorar desempenho de generalização nos usamos um treinamento com procedimento de ruído [45]. Nos criamos padrões de treinamento virtuais ("ruidoso"), adicionando ruído médio de branco Gaussiano zero, na SNR de 20 dB para o conjunto de características globais Ψ. Desta maneira para cada conjunto de características globais nos criamos quatro conjuntos virtuais, e o treinamento é baseada na união das características "originais" e "ruidosas".

APÊNDICE V

AVALIAÇÃO DE DESEMPENHO

Este apêndice apresenta resultados dos experimentos, com relação a ambos, precisão de prognóstico e complexidade computacional do método proposto. O desempenho do método proposto é comparado ao método padronizado da ITU-T P.563. O desempenho da estimativa é avaliado usando um coeficiente R de correlação por condição entre a qualidade prognosticada $\hat{Q}$ e a qualidade subjetiva Q de acordo com a equação:

$$R = \frac{\sum_i (\hat{Q}_i - \mu_{\hat{Q}})(Q_i - \mu_Q)}{\sqrt{\sum_i (\hat{Q}_i - \mu_{\hat{Q}})^2 \sum_i (Q_i - \mu_Q)^2}} \qquad (31)$$

onde μ_Q e $\mu_{\hat{Q}}$, são os valores médios das variáveis introduzidas, e somatório é sobre as condições. Tabela II contém os resultados de desempenho em termos da métrica de desempenho selecionada sobre um conjunto de teste de 7 bancos de dados da ITU-T P.Supp 23. A ITU-T P.Supp 23 Exp 1 contém distorções de codificação de fala, produzidas por sete codecs de fala padrões (predominantemente usando codec de fala usando G.729 [46]) sozinhos, ou em

configuração em tandem. Na ITU-T P.Supp 23 Exp 3 o codec de fala G.729 é avaliado sob várias condições de erros de canal como apagamento de quadro, erro de bit aleatório, e ruído de fundo. O resultado dos testes, apresentado na tabela II abaixo claramente indica que o método de avaliação de qualidade proposto executa o método padronizado da ITUT P.563.

Tempo de processamento e requisitos de memória são figuras importantes de mérito para método de avaliação de qualidades. O método de acordo com a presente invenção tem requisitos de memória insignificante: a área de armazenamento temporário de 12+12 valores escalar, calculado a partir do quadro anterior e corrente é necessário (futuros quadros não são requeridos), assim como a memória para a mistura de 12 Gaussianos.

Tabela II

COEFICIENTE DE CORRELAÇÃO POR CONDIÇÃO

R

Banco de dados	Idioma	INVENÇÃO	ITU-T P.563
Exp 1 A	Francês	0,94	0,88
Exp 1 D	Japonês	0,94	0,81
Exp 1 O	Inglês	0,95	0,90
Exp 3 A	Francês	0,93	0,87
Exp 3 C	Italiano	0,95	0,83
Exp 3 D	Japonês	0,94	0,92
Exp 3 O	Inglês	0,93	0,91

Tabela III demonstra a diferença em complexidade computacional entre o método de avaliação de qualidade proposto e o método da ITU-T P.563. A comparação é entre a implementação de ANSI-C otimizada do método da ITU-T P.563 e a implementação de MATLAB® 7 da invenção, ambos executados em uma máquina de Pentium 4 em 2,8 GHz com 1 GB de RAM. O caso onde as características de entrada $\{E^e_n, T_n, f_n\}$ estão prontamente disponíveis a partir dos codecs usados na rede, é denotado NET.

TABELA III

tempo de execução (em s) para elocuções de comprimento médio de 8 s			
	Tempo de execução (em s)		
	ITU-T P.563	Invenção	Invenção (NET)_
tempo	4,63	1,24	0,01

REFERÊNCIAS

[1] ITU-T Rec. P.830, "Subjetiva desempenho assessment of telephone-band e wideband digital codecs", 1996.

[2] ITU-T Rec. P.800, "Methods for Subjetiva Determination of Transmission Quality", 1996.

[3] ITU-R Rec. BS.1534-1, "Methods for the subjective assessment of intermediate qualidade level of coding sistemas", 2005.

[4] ITU-R Rec. BS.1284-1, "General Methods for the subjective assessment of sound qualidade ", 2003.

[5] ITU-T Rec. G.107, "The e-modelo, a computational modelo para use in transmissão planning", 2005.

[6] M. Goldstein, "Classification of methods used for assessment of text-to- speech sistemas according to the demands placed on the listener", Speech Comunicação, vol. 16, pp. 225-244, 1995.

[7] S. Quackenbush, T. Barnwell, e M. Clements, Objective Measures of Speech Qualidade. Prentice Hall, 1988.

[8] S. Wang, A. Sechave, e A. Gersho, "An objetive measure for predicting subjective quality of speech coders",IEEE J. Selected Areas in Commun., vol. 10, no. 5, pp. 819-829, 1992.

[9] J. Beerends e J. Stemerdink, "A perceptual speech-quality measure based on a psychoacoustic sound representation",J. áudio Eng. Soc, vol. 42, no. 3, pp. 115-123, 1994.

[10] S. Voran, "Objetivo estimation of perceived speech quality - Part I: Development of the measuring normalizing block technique",IEEE Trans. Speech, áudio Processing, vol. 7, no. 4, pp. 371-382, 1999.

[11] S. Voran, "Objective estimation of perceived speech uality - Part II: Evaluation of the measuring normalizing block technique",IEEE Trans. Speech, áudio Processing, vol. 7, no. 4, pp. 383-390,

1999.

[12] ITU-T Rec. P. 862, "Perceptual evaluation of speech quality (PESQ)",2001.

[13] ITU-R. BS.1387-1, "Method for Objective Measurements of Perceived Audio Quality (PEAQ)", 2001.

[14] O. Au e K. Lam, "A novel output-based objective speech quality measure para fro qireless communication",Signal Processing Proceedings, 4th Int. Conf., vol. 1, pp. 666-669, 1998.

[15] P. Gray, M. Hollier, e R. Massara, "Non-intrusive speech-quality assessment using vocal-tract modelos",em Proc. IEE Vision, Image e Signal Processing, vol. 147, pp. 493-501, 2000.

[16] J. Liang e R. Kubichek, "Output-based objective speech quality",IEEE 44th Vehicular tecnologia Conf., vol. 3, no. 8-10, pp. 1719-1723, 1994.

[17] H. Hermansky, "Perceptual linear predictiono (PLP) analysis of speech",J. Acous. Soc. Amer., vol. 87, pp. 1738-1752, 1990.

[18] T. Falk, Q. Xu, e W.-Y. Chan, "Non-intrusive GMM-based speech quality measurement",in Proc. IEEE Int. Conf. Acous., Speech, Signal Processing, vol. 1, pp. 125-128, 2005.

[19] G. Chen e V. Parsa, "Bayesian model based non-intrusive speech quality evaluation",in Proc. IEEE Int. Conf. Acous., Speech, Signal Processing, vol. 1, pp. 385-388, 2005.

[20] D. Kim, "ANIQUE: An auditory model for single-ended speech quality estimation",IEEE Trans. Speech, Audio Processing, vol. 13, pp. 821- 831, 2005.

[21] D. Kim e A. Tarraf, "Enhanced perceptual model for non-intrusive speech quality assessment", em Proc. IEEE Int. Conf. Acous., Speech, Signal Processing, vol. 1, pp. 829-832, 2006.

[22] ITU-T P. 563, "Single ended method for objective speech

quality assesment in narrow-band telephony appliations",2004.

[23] M. Werner, T. Junge, e P. Vary, "Quality control for AMR speech channels in GSM networks",in Proc. IEEE Int. Conf. Acous., Speech, Signal Processing, vol. 3, pp. 1076-1079, 2004.

[24] B. C. J. Moore, An Introduction to the Psychology of Hearing. London:

Academic Press, 1989.

[25] Z. Wang, A. Bovik, H. Sheikh, e E. Simoncelli, "Image quality assessment: From error visibility to estrutural similarity",IEEE Trans. Image Process, vol. 13, pp. 600-612, 2004.

[26] R. Reynolds e A. Rix, "Quality VoIP -an engineering challenge",BT tecnologia Journal, vol. 19, pp. 23-32, 2001.

[27] M. Schroeder e B. Atal, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates", em Proc. IEEE Int. Conf. Acous., Speech, Signal Processing, vol. 10, pp. 937-940, 1985.

[28] S. Jayant e P. Noll, Digital Coding of Waveforms. Englewood Cliffs NJ: Prentice-Hall, 1984.

[29] H. Knagenhjelm e W. B. Kleijn, "Spectral dynamics is more important than spectral distortion", em Proc. IEEE Int. Conf. Acous., Speech, Signal

Processing, vol. 1, pp. 732-735, 1995.

[30] F. Norden e T. Eriksson, "Time evolution in LPC spectrum coding",IEEE

Trans. Speech, audio Processing, vol. 12, pp. 290-301, 2004.

[31] T. Quatieri e R. Dunn, "Speech enhancement based on auditory spectral change", em Proc. IEEE Int. Conf. Acous., Speech, Signal Processing, vol. 1, pp. 257-260, 2002.

[32] J. Beauchamp, "Synthesis by spectral amplitude e brightness matching of analyzed musical instrument tones",J. audio Eng. Soc,

vol. 30, pp. 396- 406, 1982.

[33] S. Voran, "A simplified versão of the ITU algoritmo for objective measurement of speech codec quality", em Proc. IEEE Int. Conf. Acous., Speech, Signal Processing, vol. 1, pp. 537-540, 1998.

[34] P. Devijver e J. Kittler, Pattern Recognition: A Statistical Approach. London, UK: Prentice Hall, 1982.

[35] S. Stearns, "On selecting features for pattern classifiers", em Proc. 3rd Int.Conf. on Pattern Recognition, pp. 71-75, 1976.

[36] P. Pudil, F. Ferri, J. Novovicova, e J. Kittler, "Floating Search methods for featue selection with nonmonotonic criterion functions", em Proc. IEEE Intl.Conf. Pattern Recognition, pp. 279-283, 1994.

[37] T. Soderstrom, Discrete-time Stochastic Sistemas. London: Springer-Verlag, segunda ed., 2002.

[38] A. Dempster, N. Lair, e D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm",Journal Royal Statistical Society., vol. 39, pp. 1- 38, 1977.

[39] S. M. Kay, Fundamentals of Statistical Signal Processing, estimation Theory.Prentice Hall, 1993.

[40] W. B. Kleijn, P. Kroon, L. Célulaario, e D. Sereno, "A 5.85 kbps CELP algorithm for cellular applications", em Proc. IEEE Int. Conf. Acous., Speech,

Signal Processing, vol. 2, pp. 596-599, 1993.

[41] R. Laroia, N. Phamdo, e N. Farvardin, "Robust e efficient quantization of speech LSP parameters using structured vetor quantizers", em Proc. IEEE Int. Conf. Acous., Speech, Signal Processing, vol. 1, pp. 641- 644, 1991.

[42] DARPA-TIMIT, "Acoustic-telefonetic continuous speech corpus, NIST Speech Disc 1-1.1",1990.

[43] ITU-T Rec. P. Supplement 23, "ITU-T coded-speech

database",1998.

[44] ITU-T. Rec. P.810, "Modulated Noise Reference Unit",1996.

[45] R. Duda, P. Hart, e D. Stork, Pattern Classification. Wiley-Interscience, segunda ed., 2001.

[46] ITU-T. Rec. G.729, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear predicition (CS-ACELP)",1996.

REIVINDICAÇÕES

1. Método de avaliação de qualidade de sinal não intrusivo, caracterizado pelo fato de incluir os estágios de:

- determinar os parâmetros $\{E^e_n, T_n, f_n\}$ representando quadros de um sinal;

- extrair uma coleção de vetores de característica por quadro(Φ(n)) representando a informação estrutural dos quadros selecionados($\tilde{\Omega}$) do mencionado sinal a partir dos mencionados parâmetros;

- determinar um conjunto de características global ($\tilde{\Psi}$) sobre mencionada coleção de vetores de características (Φ (n)) a partir dos momentos estatísticos predeterminados dos componentes do vetor de características selecionado (Φ_1, Φ_2,...Φ_{11});

- fazer prognóstico de uma medida de qualidade de sinal ($\tilde{Q}$) a partir do mencionado conjunto de características globais ($\tilde{\Psi}$).

2. Método de acordo com a reivindicação 1, caracterizado pelo fato de incluir o estágio de selecionar somente quadros ($\tilde{\Omega}$) com um vetor de características (Φ (n)) ficando dentro de uma janela multidimensional predeterminada (Θ).

3. Método de acordo com a reivindicação 1 ou 2, caracterizado pelo fato de incluir o estágio de fazer prognóstico da mencionada medida de qualidade de sinal através de mapeamento do modelo de probabilidade de mistura Gaussiano.

4. Método de acordo com qualquer das reivindicações anteriores, caracterizado pelo fato de incluir o estágio de determinar mencionado conjunto de características global ($\tilde{\Psi}$) a partir de, pelo menos, algumas das propriedades estatísticas média, variância, inclinação e kurtosis dos componentes do mencionado vetor de características selecionado.

5. Método de acordo com a reivindicação 4, caracterizado pelo fato de incluir o estágio de determinar as propriedades estatísticas dos

momentos centrais predeterminados (μ_Φ, μ_Φ^2, μ_Φ^3 e μ_Φ^4) dos componentes do mencionado vetor de características selecionado.

6. Método de acordo com a reivindicação 5, caracterizado pelo fato de incluir o estágio de, de forma recursiva, determinar mencionados momentos centrais predeterminados (μ_Φ, $\mu_{\Phi 2}$, $\mu_{\Phi 3}$ e $\mu_{\Phi 4}$).

7. Método de acordo com qualquer das reivindicações anteriores, caracterizado pelo fato de incluir o estágio de obter mencionados parâmetros a partir de uma seqüência de bit representando mencionado sinal.

8. Método de acordo com qualquer das reivindicações anteriores 1 a 6, caracterizado pelo fato de incluir o estágio de obter mencionados parâmetros a partir da forma de onda do mencionado sinal.

9. Método de acordo com qualquer das reivindicações anteriores, caracterizado pelo fato de que mencionado sinal é um sinal de fala.

10. Método de acordo com qualquer das reivindicações anteriores, caracterizado pelo fato de que mencionado vetor de características inclui, pelo menos, algumas das características: planicidade espectral (Φ_1), dinâmica espectral (Φ_2), centróide espectral (Φ_3), variância de excitação (Φ_4), variância de sinal (Φ_5), período de ocorrência (Φ_6), e suas derivadas no tempo ($\Phi_7 - \Phi_{11}$).

11. Aparelho de avaliação de qualidade de sinal não intrusivo, caracterizado pelo fato de incluir:

- um calculador de vetor de características (60) para determinar parâmetros $\{E^e_n, T_n, f_n\}$ representando quadros (Ω) de um sinal e extraindo vetores de características por quadro ($\Phi(n)$) representando informação estrutural do mencionado sinal a partir dos mencionados parâmetros;

- um seletor de quadro (62) para selecionar uma coleção de vetores de características por quadro (Φ (n));

- meios (66, 68, 70, 72, 74) para determinar um conjunto de

características globais ($\tilde{\Psi}$) sobre mencionada coleção de vetores de características por quadro ($\Phi(n)$) a partir dos momentos estatísticos predeterminados dos componentes do vetor de características selecionado (Φ_1, Φ_2,...Φ_{11});

- um previsor de qualidade para fazer prognóstico de uma medida de qualidade de sinal ($\tilde{Q}$) a partir do mencionado conjunto de características globais ($\tilde{\Psi}$).

12. Aparelho de acordo com a reivindicação 11, caracterizado pelo fato de que mencionado seletor de quadro (62) é arranjado para incluir somente quadros ($\tilde{\Omega}$) com um vetor de características (Φ (n)) ficando dentro de uma janela multidimensional predeterminada (Θ) na mencionada coleção.

13. Aparelho de acordo com a reivindicação 11 ou 12, caracterizado pelo fato de que mencionado previsor de qualidade é arranjado fazer prognóstico da mencionada medida de qualidade do sinal através de mapeamento do modelo de probabilidade de mistura Gaussiano.

14. Aparelho de acordo com a reivindicação 11, 12 ou 13, caracterizado pelo fato de que mencionados meios (66, 68, 70, 72, 74) para determinar mencionado conjunto de características globais ($\tilde{\Psi}$) é arranjado para determinar, pelo menos, algumas das propriedades estatísticas médias, variância, inclinação e kurtosis dos componentes do mencionado vetor de características selecionado.

15. Aparelho de acordo com a reivindicação 14, caracterizado pelo fato de que mencionados meios (66, 68, 70, 72, 74) para determinar mencionado conjunto de características globais (Ψ) é arranjado para determinar as propriedades estatísticas dos momentos centrais predeterminados (μ_{Φ}, $\mu_{\Phi 2}$, $\mu_{\Phi 3}$ e $\mu_{\Phi 4}$) dos componentes do mencionado vetor de características selecionado.

16. Aparelho de acordo com a reivindicação 15, caracterizado pelo fato de que os mencionados meios (66, 68, 70, 72, 74) para determinar

mencionado conjunto de características globais ($\tilde{\Psi}$) são arranjados para, de forma recursiva, determinar mencionados momentos centrais predeterminados (μ_{Φ}, $\mu_{\Phi 2}$, $\mu_{\Phi 3}$ e $\mu_{\Phi 4}$).

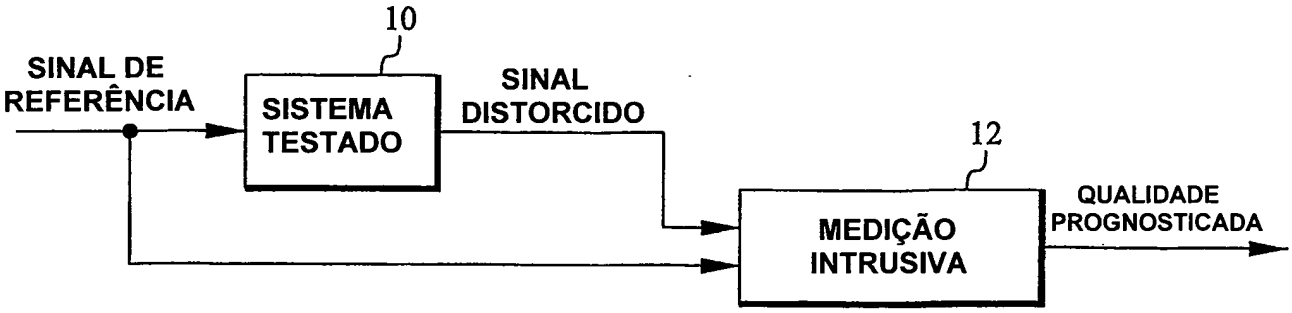


FIG. 1

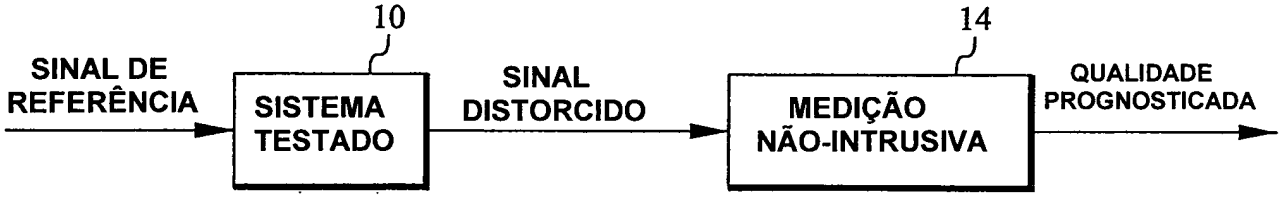


FIG. 2

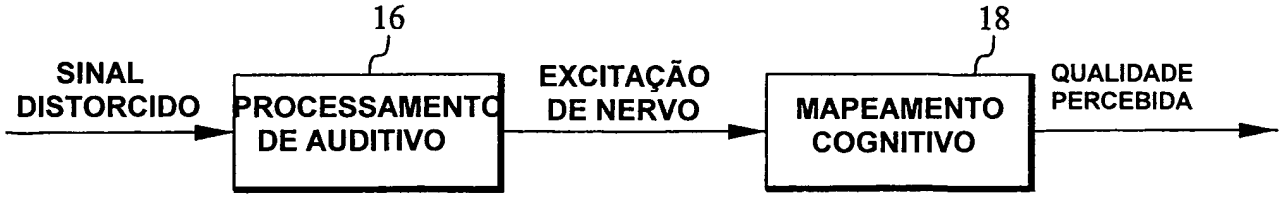


FIG. 3

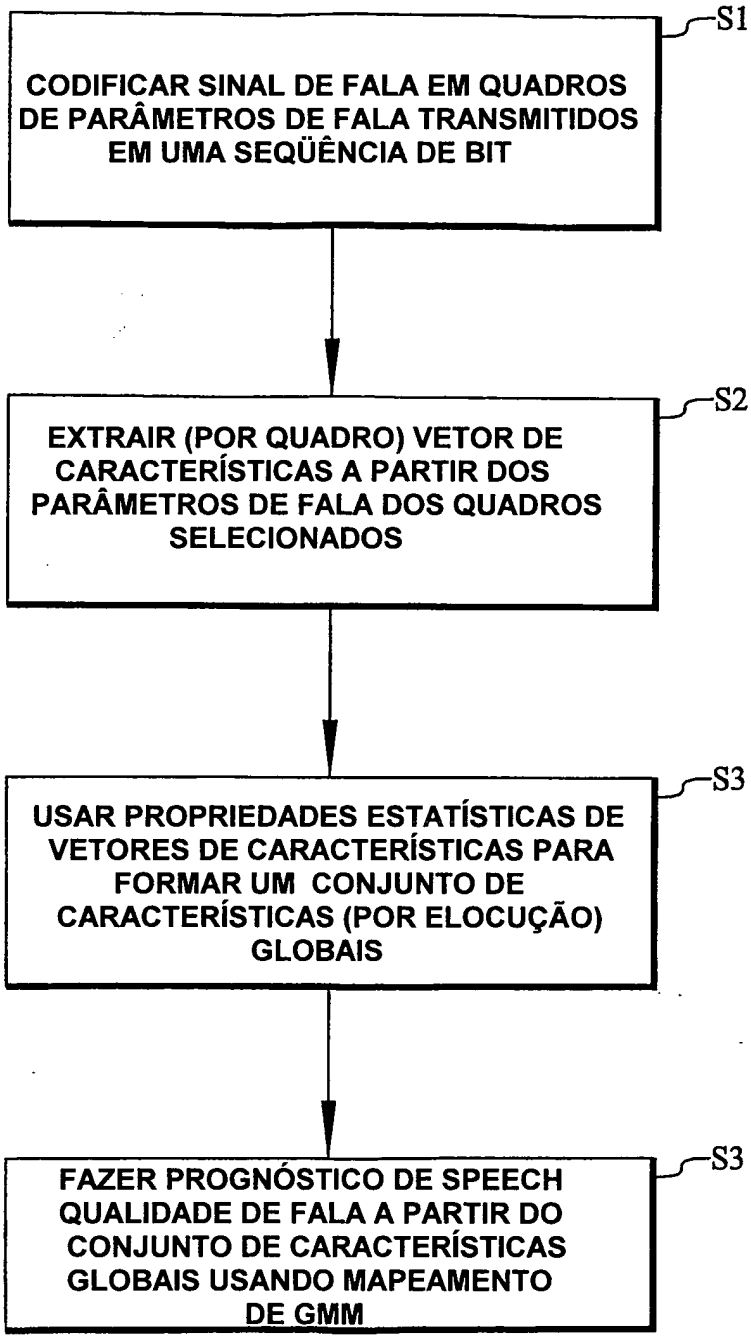

S1
CODIFICAR SINAL DE FALA EM QUADROS DE PARÂMETROS DE FALA TRANSMITIDOS EM UMA SEQÜÊNCIA DE BIT
S2
EXTRAIR (POR QUADRO) VETOR DE CARACTERÍSTICAS A PARTIR DOS PARÂMETROS DE FALA DOS QUADROS SELECIONADOS
S3
USAR PROPRIEDADES ESTATÍSTICAS DE VETORES DE CARACTERÍSTICAS PARA FORMAR UM CONJUNTO DE CARACTERÍSTICAS (POR ELOCUÇÃO) GLOBAIS
S3
FAZER PROGNÓSTICO DE SPEECH QUALIDADE DE FALA A PARTIR DO CONJUNTO DE CARACTERÍSTICAS GLOBAIS USANDO MAPEAMENTO DE GMM

FIG. 4

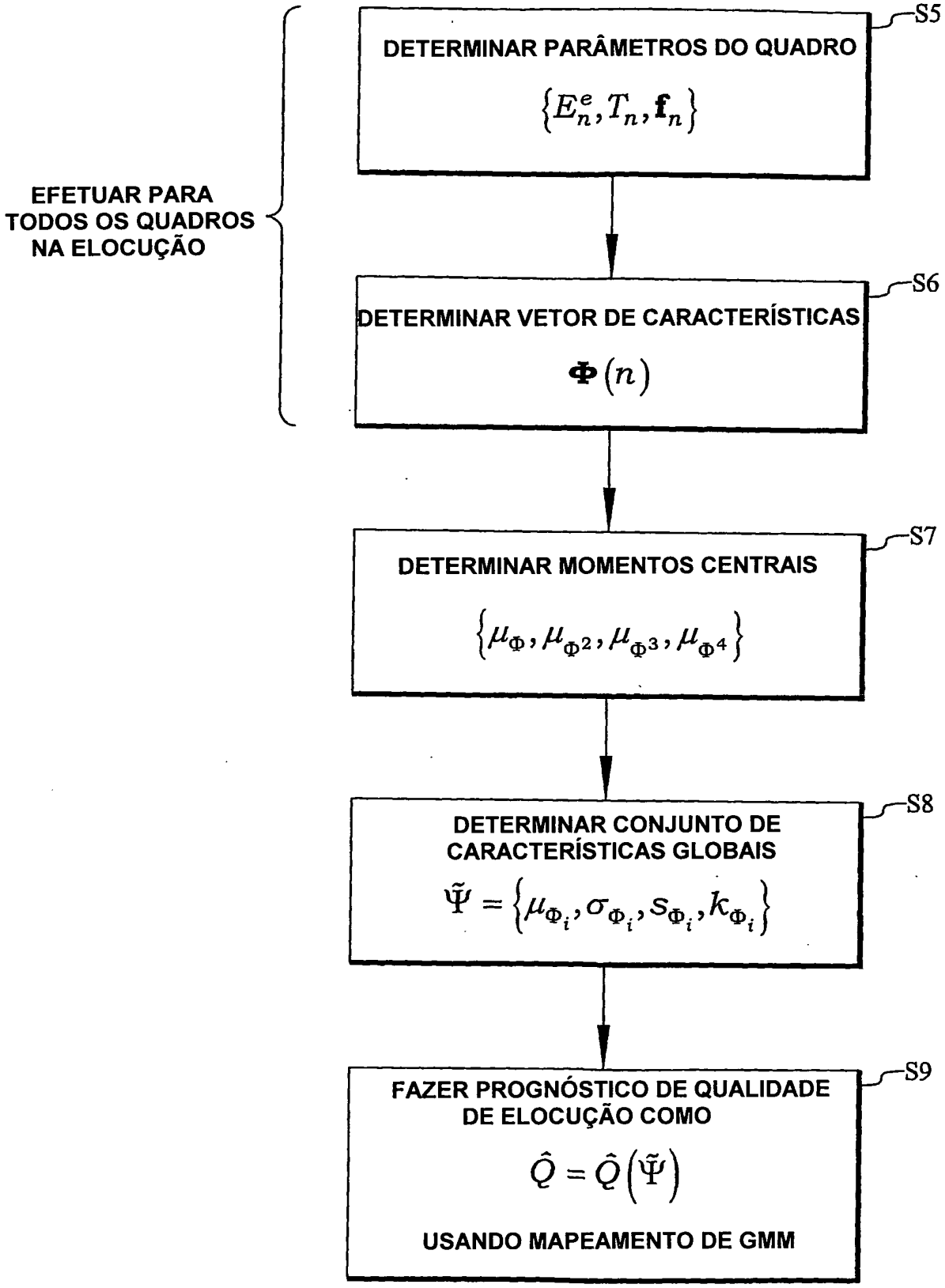

EFETUAR PARA TODOS OS QUADROS NA ELOCUÇÃO
S5
DETERMINAR PARÂMETROS DO QUADRO
$\{E_n^e, T_n, \mathbf{f}_n\}$
S6
DETERMINAR VETOR DE CARACTERÍSTICAS
$\mathbf{\Phi}(n)$
S7
DETERMINAR MOMENTOS CENTRAIS
$\{\mu_\Phi, \mu_{\Phi^2}, \mu_{\Phi^3}, \mu_{\Phi^4}\}$
S8
DETERMINAR CONJUNTO DE CARACTERÍSTICAS GLOBAIS
$\tilde{\Psi} = \{\mu_{\Phi_i}, \sigma_{\Phi_i}, s_{\Phi_i}, k_{\Phi_i}\}$
S9
FAZER PROGNÓSTICO DE QUALIDADE DE ELOCUÇÃO COMO
$\hat{Q} = \hat{Q}(\tilde{\Psi})$
USANDO MAPEAMENTO DE GMM

FIG. 5

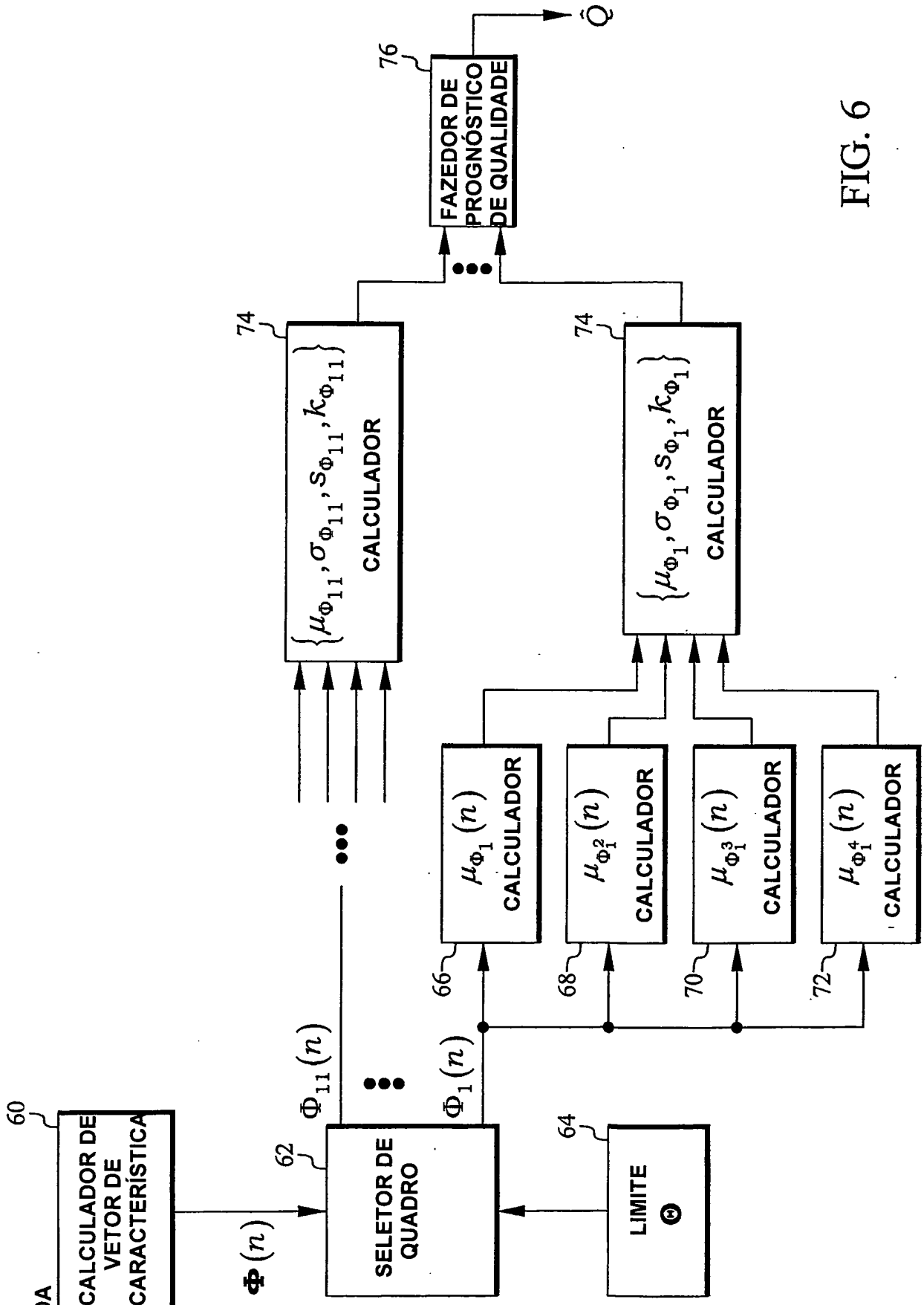

FALA CODIFICADA
60
CALCULADOR DE VETOR DE CARACTERÍSTICA
$\Phi(n)$
62
SELETOR DE QUADRO
$\Phi_{11}(n)$
$\Phi_1(n)$
64
LIMITE
Θ
66
$\mu_{\Phi_1}(n)$ CALCULADOR
68
$\mu_{\Phi_1^2}(n)$ CALCULADOR
70
$\mu_{\Phi_1^3}(n)$ CALCULADOR
72
$\mu_{\Phi_1^4}(n)$ CALCULADOR
74
$\{\mu_{\Phi_{11}}, \sigma_{\Phi_{11}}, s_{\Phi_{11}}, k_{\Phi_{11}}\}$ CALCULADOR
74
$\{\mu_{\Phi_1}, \sigma_{\Phi_1}, s_{\Phi_1}, k_{\Phi_1}\}$ CALCULADOR
76
FAZEDOR DE PROGNÓSTICO DE QUALIDADE
$\hat{Q}$

FIG. 6

PI0707343-7

RESUMO

"MÉTODO E APARELHO DE AVALIAÇÃO DE QUALIDADE DE SINAL NÃO INTRUSIVO"

Um aparelho de avaliação de qualidade de sinal não intrusivo inclui um calculador de vetor de característica (60) que determina parâmetros representando quadros de um sinal e extrai uma coleção de vetores de característica por quadro (Φ(n)) representando informação estrutural do sinal a partir dos parâmetros. Um seletor de quadro (62) preferencialmente seleciona somente quadros (Ω) com um vetor de característica (Φ(n)) ficando dentro de uma janela de múltiplas dimensões predeterminada (θ) . Meios (66, 68, 70, 72, 74) determinam um conjunto de características global (Ψ) sobre a coleção de vetores de características (Φ(n)) de momentos estatísticos dos componentes de vetor de característica selecionados ((1^,02,...O11) . Um previsor de qualidade (76) faz prognóstico de uma medida de qualidade de sinal (Qj do conjunto de características global (Ψ) .