

Um Modelo para Transcrição Automática de Harmonia Musical

Francisco P. C. Sant'Anna e José Mauro P. Fortes

Resumo—Uma abordagem para o projeto de um sistema capaz de transcrever seqüências de acordes em peças musicais é apresentada. O sistema proposto analisa um sinal de áudio e retorna as cifras correspondentes aos acordes que melhor representam o sinal.

Palavras-Chave—Transcrição, Música, Harmonia, Processamento de Sinais Musicais, Identificação de acordes.

Abstract—A system capable of transcribing chord sequences in musical pieces is described. The system analyzes an audio signal and returns the symbols corresponding to the chords that best fit the signal.

Keywords—Transcription, Music, Harmony, Musical signal processing, Chord identification.

I. INTRODUÇÃO

Transcrição de música é o ato de ouvir uma peça musical e escrever a notação correspondente aos seus constituintes, transformando um sinal acústico em uma representação simbólica. Esforços para a transcrição automática de música polifônica vêm sendo empreendidos desde 1970. A maior parte dos trabalhos busca somar conhecimentos e técnicas focando a transcrição integral da partitura musical, especialmente de peças eruditas, considerando em detalhes os sons concorrentes. Apesar de alguns expressivos avanços alcançados, a complexidade de tratar a superposição de sons e eventos ainda impõe severas restrições à eficiência dos resultados.

Uma informação bastante relevante, passível de ser transcrita dos sinais musicais, diz respeito à sua estrutura harmônica. Harmonia é a seqüência de acordes que ordenam os sinais musicais, atribuindo-lhes uma cor característica e uma estrutura agradável e familiar ao ouvinte. É a base melódica - um conjunto harmônico de freqüências - sobre a qual se acomoda a melodia principal. Facilmente identificada por músicos experientes, a harmonia costuma ser de difícil percepção por aqueles sem uma educação musical adequada.

Na música popular a interpretação é muito mais livre que na erudita. Por consequência disso, o registro musical é mais genérico e freqüentemente se resume à notação da harmonia acompanhada da linha melódica principal. A partitura completa é geralmente dispensada, sendo a notação apenas uma diretriz ao executor. Neste sentido, a transcrição de música popular torna-se uma tarefa muito mais simples que a da erudita (na qual é preciso detalhar cada linha melódica de cada instrumento). Pode-se dizer que a transcrição da harmonia é a essência da transcrição de música popular. O escopo deste trabalho é a transcrição automática de harmonia musical, a partir

de gravações reais. Atendo-se a esta visão mais simples, que dispensa os esforços de separação das componentes sonoras simultâneas de diferentes fontes, esperamos obter resultados efetivos e aplicáveis.

A existência de um método de transcrição de harmonia de forma automática traz benefícios não só para músicos menos experientes que continuamente precisam recorrer a livros de músicas cifradas, mas também abre um leque de alternativas nas áreas de música interativa, análise e classificação automática de arquivos musicais, entre outras diversas possibilidades imagináveis.

II. ABORDAGEM AO PROBLEMA

A aptidão para a transcrição de harmonia por um músico fundamenta-se em dois aspectos principais: um bom ouvido e o conhecimento de harmonia funcional. O conceito de “bom ouvido” refere-se à capacidade de distinguir as notas musicais em melodias, assim como intervalos entre notas simultâneas ou consecutivas. Já a harmonia funcional associa os acordes à sua função no encadeamento harmônico, isto é, se trata-se de um acorde estável, de preparação, passagem ou conclusão: qual o papel de cada acorde em cada tonalidade. A primeira habilidade permite a identificação de parte das notas e intervalos do acorde que está soando, a segunda permite que se conheçam seqüências típicas de acordes sabendo-se o que esperar da seqüência e antecipando as transições entre acordes. A condução de algumas linhas melódicas, sobretudo linhas de baixo, também fornece elementos para a identificação dos acordes.

Um modelo que busque reproduzir o processo de transcrição de harmonia por um músico treinado deve contemplar tanto o aspecto da identificação de notas e intervalos quanto a previsibilidade possibilitada pelo conhecimento de harmonia funcional. Apesar da análise sonora para a identificação de notas já ser alvo de amplos estudos, não se tem conhecimento de trabalhos que conjuguem o conhecimento a priori de seqüências típicas de acordes à análise espectral e identificação das notas.

Tendo como prioridade a modelagem da harmonia funcional e seu uso na determinação da probabilidade de seqüências de acordes, utilizamos modelos simplificados de identificação de notas, sem nos ater a métodos de processamento de sinais que busquem aperfeiçoar pontualmente esta capacidade. Inicialmente consideramos que a tonalidade do trecho musical a ser transcrito é conhecida a priori.

A abordagem aqui sugerida para a transcrição da harmonia de uma peça musical consiste em amostrar o sinal de áudio e segmentá-lo em janelas de tempo, nas quais será feita a

Francisco P. C. Sant'Anna e José Mauro P. Fortes, CETUC (Centro de Estudos de Telecomunicações da Universidade Católica), Pontifícia Universidade Católica do Rio de Janeiro, RJ, Brasil. E-mails: fpsantanna@oi.com.br, jmfortes@cetuc.puc-rio.br.

com $j = 1, \dots, N$, sendo as probabilidades de transição (de 1 passo) entre estados dada por

$$q_{ij}(t) = P(s_{t+1} = j \mid s_t = i) \quad (2)$$

2) *Condição de Estacionariedade*: As probabilidades de transição e de estados são invariantes no tempo, ou seja, as probabilidades de transição (de 1 passo) $q_{ij}(t)$ e as probabilidades de estado $p_j(t)$ não dependem de t . Assim,

$$P(s_{t_1+1} = j \mid s_{t_1} = i) = P(s_{t_2+1} = j \mid s_{t_2} = i) = q_{ij} \quad (3)$$

e

$$P(s_{t_1} = j) = P(s_{t_2} = j) = p_j \quad (4)$$

3) *Condição de Independência das Saídas*: Os vetores aleatórios \mathbf{x} correspondentes às saídas (observações) dos estados são estatisticamente independentes:

$$p_{\mathbf{x}_1, \dots, \mathbf{x}_T}(\mathbf{x}_1, \dots, \mathbf{x}_T \mid s_1, \dots, s_T) = \prod_{t=1}^T p_{\mathbf{x}_t}(\mathbf{x}_t \mid s_t) \quad (5)$$

Tanto a condição de estacionariedade quanto a de independência entre as saídas não são totalmente verificadas em nosso objeto de estudo, no entanto assumí-las não implica em grandes distorções e simplifica consideravelmente o tratamento matemático do problema.

V. PARAMETRIZAÇÃO DO HMM APLICADO AO PROBLEMA

Para modelar as seqüências de harmonia musical a partir de Cadeias de Markov Escondidas, há que se definir o relacionamento que mapeia os acordes em um conjunto de estados, as realizações sonoras desses acordes nas saídas e as mudanças de acorde nas transições de estados.

A. Mapeamento de Acordes com Estados

Dentre as possíveis parametrizações do modelo, a mais direta é aquela em que cada acorde deve ser representado por um estado. Isto significa que se tomarmos acordes com nome em cada um dos 12 graus, cada qual podendo ser maior, menor, ou diminuto, teremos $12 \times 3 = 36$ estados. Se considerarmos que cada um desses acordes pode possuir sétima maior, menor ou não apresentá-la, já são $36 \times 3 = 108$ estados. E a cada tensão que se adiciona no conjunto de possibilidades e que possa ser omitida ou aparecer como maior ou menor (alternativamente como aumentada ou diminuta, dependendo do grau) multiplicamos por 3 o total de estados. Para uma modelagem consideravelmente completa, com 6^a , 9^a , 11^a e 13^a teríamos $3^4 \times 108 = 8748$ estados.

Evidentemente grande parte desses acordes seria bastante rara, podendo ser aproximada por acordes parecidos. Vê-se portanto que mais importante - e certamente mais eficiente - que criar um conjunto completo de acordes é construir um conjunto bem representativo destes.

A idéia de construir um conjunto incompleto de estados, que represente de forma eficiente todos os possíveis acordes pode ser implementada de forma otimizada através da adaptação de um algoritmo de treinamento e criação de dicionário de quantização vetorial como o LBG [2]. Neste tipo de algoritmo buscam-se os N pontos que, ao serem utilizados

na representação do conjunto total de pontos, minimizam a distorção. Para prosseguir neste sentido seria necessário definir o critério de distorção entre acordes. Embora este conceito possa ser de extrema relevância neste contexto de análise de harmonia musical, optou-se por não tratá-lo neste momento. Antes de estudar a forma com que o número de estados afeta o desempenho do modelo, estarão sendo usados modelos de estados intuitivos, com níveis de complexidade harmônica progressiva (cada vez mais acordes), que sejam considerados (subjetivamente) os mais representativos dos repertórios a serem tratados. Esta análise subjetiva de representatividade corresponde a selecionar os acordes para o modelo a partir do estudo da frequência relativa de aparição destes no repertório, combinada com a verificação de substitutos próximos, que possam representá-los com pouca distorção. Este é um conceito bastante próximo ao do LBG, apenas com uma definição intuitiva de distorção.

Partimos portanto de modelos básicos, com não mais que 6 acordes (que podem representar muito bem um repertório restrito), aumentando seu nível de complexidade para espaços mais complexos à medida que avançarmos nos resultados.

B. Mapeamento das Saídas

Conforme apresentado na seção III, as saídas de cada janela de tempo são representadas através de vetores de 12 posições, uma para cada nota, cada qual com o nível espectral de energia médio da nota nas oitavas de interesse. Estes vetores são normalizados, tornando os resultados independentes do volume (intensidade total do som) de cada trecho. As densidades de probabilidade de saída de cada estado $p_{\mathbf{x}}(\mathbf{X})$ são portanto funções de $\mathbf{R}^{12} \rightarrow \mathbf{R}$.

A forma mais intuitiva de treinar estas densidades seria, a partir das observações de uma seqüência de treinamento \mathbf{T} , dividir o espaço de 12 dimensões em regiões e verificar a frequência de ocorrência de saídas do estado em cada região. Ocorre que o número de regiões do espaço 12-dimensional fica invariavelmente extenso demais: mesmo que utilizemos apenas 2 níveis de intensidade por dimensão - o que certamente seria muito impreciso - já teríamos $2^{12} = 4.096$ regiões. O número elevado de regiões inviabiliza o treinamento por frequência relativa de ocorrência.

Alternativamente, buscou-se efetuar uma análise estatística das ocorrências das saídas dos estados, objetivando identificar aproximações com densidades conhecidas. A aproximação gaussiana mostrou-se bastante aplicável à modelagem, sobretudo à medida que aumentam o número de amostras de cada estado. Nesta aproximação, dado o estado (estado j por exemplo), o vetor de saída do estado \mathbf{x} é considerado gaussiano, com média $\mathbf{m}^{(j)}$ e matriz covariância $\mathbf{K}^{(j)}$. Note que, sob a hipótese gaussiana, a densidade de probabilidade da saída de cada um dos possíveis estados fica completamente definida a partir de seu vetor média e sua matriz covariância:

$$p_{\mathbf{x}|j}(\mathbf{X}) \rightarrow N(\mathbf{m}^{(j)}, \mathbf{K}^{(j)}) \quad (6)$$

A título de ilustração, a figura 2 mostra como a aproximação gaussiana é adequada.

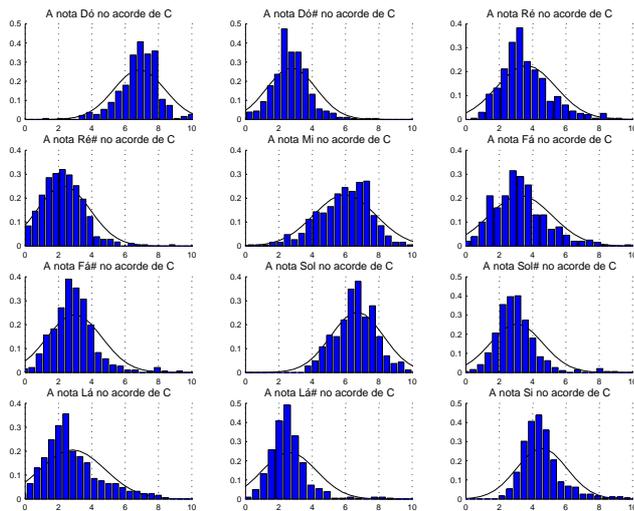


Fig. 2. Histograma e aproximação gaussiana das componentes da saída do estado correspondente ao acorde dó maior

C. Mapeamento das Transições de Estados

As transições de estados podem ser tratadas de duas formas diferentes: pode-se considerar que ocorre uma transição a cada mudança de janela de amostragem, ou podem-se chamar transições apenas aos eventos de mudança de acorde.

A primeira alternativa tem a grande vantagem da padronização do intervalo entre transições, que seria igual à duração de uma janela de amostragem. No entanto, como em geral um acorde estende-se por diversas janelas de amostragem, a maior parte das transições seria de cada estado para si mesmo (sem haver mudança de acorde). Isso gera um desalinhamento entre transições de janelas de amostragem e transição de acordes, de modo que o número de janelas por acorde no trecho a ser transcrito passa a interferir fortemente no resultado. Trechos com ritmo harmônico mais lento (e portanto com mais janelas por acorde) acabam tendo probabilidades de auto-transição (transição de um estado para si mesmo) mais altas do que aquelas que mudam mais rápido de acorde. Como no sistema real não há correlação entre o ritmo harmônico e a seqüência harmônica, não é recomendável que o modelo crie este tipo de vínculo. Temos portanto uma alternativa com a vantagem da fácil identificação das transições no tempo e a desvantagem das distorções nas probabilidades de transição de estados e vinculação destas ao ritmo harmônico.

A segunda alternativa considera como transições apenas as passagens de janela de amostragem onde há mudança de acordes. Isto traz total alinhamento com a identidade estado=acorde, fazendo com que a probabilidade de uma seqüência de transições de estados seja efetivamente a probabilidade da sucessão de acordes correspondentes, independente da duração de cada acorde. No entanto esta solução elimina a periodicidade regular das transições, criando um novo problema: estimar os instantes em que ocorrem as mudanças de acorde.

Ambas as alternativas foram estudadas. Optamos por trabalhar inicialmente com a primeira abordagem, não tendo se identificado significativo ganho de performance ao avançar na

estimativa dos instantes das transições.

VI. TREINAMENTO

Os parâmetros $\mathbf{m}^{(j)}$ e $\mathbf{K}^{(j)}$, que definem as densidades de saída de cada estado j , $1 \leq j \leq N$, assim como as probabilidades de transição q_{ij} e de estado p_j , devem ser estimados a partir de um processo de treinamento do HMM proposto para o sistema.

A. Definição da Seqüência de Treinamento

Para treinar o HMM utilizamos uma seqüência de treinamento \mathcal{T} . Esta seqüência é obtida a partir da concatenação de diversos trechos musicais, gerando um conjunto de amostras que são pares de vetores saídas com o estado que o gerou. Se a duração total dos trechos musicais utilizados na seqüência de treinamento for T e o tamanho da janela de amostragem for w , então \mathcal{T} apresentará $N_s = \frac{T}{w}$ amostras, sendo a k -ésima amostra o par (\mathbf{X}_k, s_k) ; $\mathbf{X}_k \in \mathbf{R}^{12}$ e $s_k \in \{1 \dots N\}$.

$$\mathcal{T} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 & \dots & \mathbf{X}_{N_s} \\ s_1 & s_2 & \dots & s_{N_s} \end{bmatrix} \quad (7)$$

A seqüência \mathcal{T} deve ser representativa do grupo de sinais a ser transcrito. Na aplicação em questão, esta representatividade se reflete principalmente na natureza da execução e na familiaridade do repertório.

A natureza da execução diz respeito às características de realização dos acordes: quais instrumentos estão envolvidos, características de gravação e os elementos que formam o espectro típico de cada acorde. Para que as densidades de probabilidade de saída sejam as mais próximas daquelas que geram os sinais a serem analisados na transcrição de acordes, é extremamente importante que a seqüência de treinamento inclua sinais com a mesma natureza de execução dos sinais a serem transcritos.

O outro aspecto de destaque na representatividade da seqüência de treinamento é a familiaridade do repertório. Existe uma identidade cultural entre as peças de um mesmo repertório que implica uma maior proximidade musical, um fator de estética que acarreta semelhanças rítmicas, melódicas e harmônicas. Estilos musicais têm peculiaridades harmônicas que definem seqüências típicas traduzidas em probabilidades de transição. Portanto, músicas de um mesmo repertório tendem a apresentar probabilidades de transição de acordes mais padronizadas.

B. Treinamento das Densidades de Saída dos Estados

A densidade de saída de cada estado j , $1 \leq j \leq N$, é a gaussiana de média $\hat{\mathbf{m}}^{(j)}$ e variância $\hat{\mathbf{K}}^{(j)}$ estimadas a partir do conjunto $\{\mathbf{X}_k \in \mathcal{T} \mid s_k = j\}$; $1 \leq k \leq N_s$, contendo todas as saídas de \mathcal{T} correspondentes às N_j ocorrências do estado j na seqüência de treinamento. Os parâmetros $\hat{\mathbf{m}}^{(j)}$ e $\hat{\mathbf{K}}^{(j)}$ são os estimadores de máxima verossimilhança obtidos por

$$\hat{\mathbf{m}}^{(j)} = \frac{1}{N_j} \sum_{\ell=1}^{N_j} \mathbf{x}_\ell^{(j)} \quad (8)$$

onde $\mathbf{X}_\ell^{(j)}$ é a saída correspondente à ℓ -ésima ocorrência do estado j , além de

$$\hat{k}_{nm}^{(j)} = \frac{1}{N_j - 1} \sum_{\ell=1}^{N_j} (X_{\ell,n}^{(j)} - \hat{m}_n^{(j)})(X_{\ell,m}^{(j)} - \hat{m}_m^{(j)}) \quad (9)$$

onde $\hat{k}_{nm}^{(j)}$ ($1 \leq n \leq 12; 1 \leq m \leq 12$) são os elementos da matriz $\hat{\mathbf{K}}^{(j)}$ e $X_{\ell,n}^{(j)}$ e $\hat{m}_n^{(j)}$ representam, respectivamente, as n -ésimas componentes dos vetores $\mathbf{X}_\ell^{(j)}$ e $\hat{\mathbf{m}}^{(j)}$ que aparecem em (8).

C. Treinamento das Probabilidades de Estado e Transição

A distribuição de probabilidades dos estados do modelo, $\{p_j = P(s_t = j); j = 1, 2, \dots, N\}$, pode ser facilmente estimada a partir da frequência relativa de ocorrência de cada estado na seqüência \mathcal{T} :

$$\hat{p}_j = \frac{N_j}{N_s} \quad (10)$$

As probabilidades de transição de estado formam a matriz $\hat{\mathbf{Q}}$, com dimensões $N \times N$. Seus elementos \hat{q}_{ij} podem ser obtidos por:

$$\hat{q}_{ij} = \frac{N_{i \rightarrow j}}{N_i} \quad (11)$$

onde $N_{i \rightarrow j}$ é o número de vezes em que ocorre a transição $i \rightarrow j$ na seqüência de treinamento.

VII. IDENTIFICAÇÃO DA SEQÜÊNCIA DE ACORDES

No modelo proposto, a transcrição dos acordes de uma peça musical é feita escolhendo-se a seqüência de estados que mais provavelmente gerou as saídas observadas, isto é, a seqüência de máxima verossimilhança. Escolhe-se a seqüência de estados mais provável dado o conjunto de saídas observadas $\mathbf{X} = \{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_L\}$ e as informações a priori \hat{p}_j , \hat{q}_{ij} , $\mathbf{m}^{(j)}$ e $\mathbf{K}^{(j)}$, obtidas a partir da seqüência de treinamento. A transcrição da harmonia em um trecho com L acordes corresponde à escolha da seqüência de estados \mathcal{S}_k , $k = 1, \dots, N^L$, que corresponde à melhor dentre as N^L possíveis seqüências de acordes. Portanto, a decisão da melhor seqüência de estados $\hat{\mathcal{S}}$ deve ser tal que:

$$\hat{\mathcal{S}} = \mathcal{S}_{k^*} \quad (12)$$

onde

$$k^* = \max_k^{-1} P(\mathcal{S}_k | \mathcal{X} = \mathcal{X}) \quad (13)$$

e, conseqüentemente,

$$k^* = \max_k^{-1} p_{\mathcal{X}|\mathcal{S}_k}(\mathcal{X})P(\mathcal{S}_k) \quad (14)$$

onde \max_k^{-1} designa o valor de k que maximiza o argumento.

Esta maximização pode ser feita de forma bastante eficiente utilizando-se o algoritmo de Viterbi[3].

VIII. RESULTADOS E SIMULAÇÕES PRELIMINARES

As simulações foram realizadas em espaços de 8 estados, correspondendo a 8 acordes básicos. Os trechos musicais foram selecionados de peças de harmonia simples, representáveis de forma satisfatória pelos 8 acordes do modelo. Utilizamos duas naturezas de execuções distintas: gravações de piano isolado e execuções de bandas completas, com bateria. Utilizamos 13 trechos distintos, de tamanhos variáveis, sempre 12 trechos na seqüência de treinamento e o remanescente como peça a ser transcrita. O critério de performance foi a taxa de erro de estado, correspondendo à fração das janelas de amostragem que apresentaram erro de transcrição.

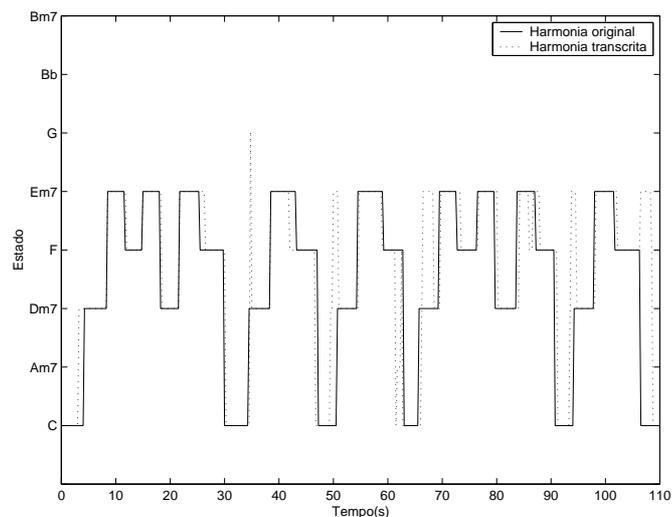


Fig. 3. Ilustração do resultado obtido na transcrição de uma peça de piano. A taxa de erro observada foi de 0,12.

A figura 3 ilustra o resultado obtido na transcrição de uma peça de piano bastante simples (apenas 5 acordes) a partir do modelo proposto. As áreas de descolamento da linha sólida com a pontilhada correspondem às regiões de erro de transcrição. Percebe-se que a maior parcela dos erros acontece nos momentos de mudança de acorde e que em geral o erro é transiente, havendo rápida convergência para o acorde correto.

Em uma avaliação preliminar, buscamos avaliar o impacto no desempenho da transcrição do que chamamos de natureza de execução. Para tanto, geramos 3 seqüências de treinamento: uma apenas com trechos de gravações de piano, outra apenas com gravações de banda (bateria, guitarras, baixo, vocais, teclados e sopros) e uma terceira com os dois tipos de sinais. Simulamos a transcrição de gravações de piano a partir de cada uma das 3 diferentes seqüências de treinamento, verificando a grande variação nos resultados, conforme apresentado na tabela I.

A transcrição de trechos de piano puro a partir do treinamento afim apresentou taxa de erro média de 6%, enquanto que quando é feito o treinamento somente com gravações de bandas, a média de erro chegou a 60%. Esta diferença mostra a grande influência da similaridade da natureza de execução no resultado. Por outro lado, ao verificarmos que a média de erro obtida a partir do treinamento com ambos os tipos de sinais ficou abaixo de 13%, podemos perceber que não há extremo

TABELA I

TAXA DE ERRO NA TRANSCRIÇÃO DE PIANO A PARTIR DE DIFERENTES T

	Seqüência de treinamento T		
	Piano	Banda	Ambos
Peça1	0,01	0,58	0,04
Peça2	0,01	0,71	0,04
Peça3	0,03	0,48	0,16
Peça4	0,07	0,43	0,09
Peça5	0,12	0,81	0,23
Peça6	0,14	0,56	0,17
Média	0,06	0,60	0,12

rigor quanto ao pleno alinhamento da natureza de execução da peça a ser transcrita com a seqüência de treinamento, bastando que a referência apresente ao menos alguns poucos representantes da categoria do sinal a ser transcrito.

Analizamos também o efeito do tamanho da janela de amostragem sobre o desempenho do sistema, conforme apresentado na tabela II. Verificam-se ganhos marginais decres-

TABELA II

TAXA DE ERRO POR TAMANHO DE JANELA

	Tamanho da janela de amostragem			
	185ms	250ms	300ms	350ms
Banda1	0,53	0,32	0,32	0,27
Banda2	0,14	0,05	0,03	0,16
Banda3	0,24	0,28	0,25	0,23
Banda4	0,11	0,18	0,15	0,17
Banda5	0,13	0,15	0,17	0,14
Banda6	0,34	0,46	0,43	0,35
Piano1	0,12	0,10	0,10	0,10
Piano2	0,21	0,18	0,18	0,18
Piano3	0,20	0,17	0,16	0,13
Piano4	0,06	0,04	0,06	0,05
Piano5	0,27	0,23	0,21	0,19
Piano6	0,01	0,04	0,05	0,05
Média	0,20	0,19	0,18	0,18

centes no desempenho ao se aumentar o tamanho da janela de amostragem. Do ponto de vista de implementação prática, uma janela de amostragem de 250 ms apresenta uma adequada relação custo/benefício. Este ganho é uma evidência de problemas de resolução em baixas freqüências e vazamento de freqüência, questões de processamento de sinais que podem ser otimizadas pontualmente.

IX. CONCLUSÕES

Os resultados preliminares obtidos indicam a pertinência do modelo e a viabilidade de implementação do objetivo proposto.

O crescimento do espaço de acordes e a evolução para seqüências harmônicas e naturezas de execução mais diversas prejudica consideravelmente o desempenho de transcrição, mas diversos caminhos para aperfeiçoamento e possibilidades de refinamento se apresentam para a superação dos obstáculos. Inicialmente será necessário ajustar o processamento dos sinais através de técnicas que otimizem a análise espectral. Há ainda uma série de alternativas de inserir capacidade de aprendizado ao modelo - seja nas densidades de saída ou nas probabilidades de transição - deixando-o menos dependente da familiaridade

de repertório com a seqüência de treinamento. Isto permitiria que adaptativamente o modelo se adequasse à peça em análise.

Uma vantagem adicional da forma como o modelo foi concebido é a sua aplicabilidade à transcrição em tempo real. Determinando-se uma profundidade de decisão para a treliça, pode-se proceder a uma decisão sub-ótima com pequeno atraso em relação ao sinal em execução. Também a carga computacional para o processamento não é pesada, não constituindo impedimento para o processamento em tempo real. Este aspecto abre espaço para a interatividade musical com um grande potencial de aplicações.

REFERÊNCIAS

- [1] Klapuri, Virtanen, Eronen and Seppänen. "Automatic transcription of musical recordings", *Consistent and Reliable Acoustic Cues Workshop - CRAC/01*, Aalborg, Denmark, September 2001.
- [2] Y. Linde; A. Buzo and R.M. Gray, "An algorithm for vector quantizer design", *IEEE Transactions on Communications*, vol. 28, p. 84-95, January, 1980.
- [3] G.D. Forney, "The Viterbi algorithm", *Proceedings of the IEEE*, 61(3):268-278, March 1973.
- [4] L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE*, vol.77(2), p. 257-286, February 1989.