

Uma proposta do uso de adaptatividade para busca de padrões musicais

(5 de janeiro 2009)

P. R. N. Pedruzzi, R. A. Redder Jr. e J. J. Neto

Resumo — Este artigo apresenta uma proposta de utilização de adaptatividade para um sistema de busca musical. O trabalho¹ consiste na construção de um sistema capaz de identificar uma música, dentre um repertório previamente estabelecido, a partir de uma amostra de áudio contendo a execução de um pequeno trecho de uma de suas melodias, cantado ou assobiado.

O principal foco do trabalho está no reconhecimento de melodias similares que podem apresentar transposição de tonalidade, dilatação de durações, erros de notas etc.

Primeiramente são apresentados alguns detalhes sobre o projeto desenvolvido, em seguida define-se formalmente a noção de proximidade/semelhança entre melodias e, por fim, propõe-se um algoritmo para reconhecimento de melodias próximas, utilizando adaptatividade.

Palavras chave— Reconhecimento de padrões (*pattern matching*), reconhecimento de cadeias (*string matching*), melodia (*melody*), autômato (*automaton*), adaptatividade (*adaptivity*), mínimos quadrados (*least squares*).

I. INTRODUÇÃO

Este artigo apresenta os avanços obtidos no trabalho de um projeto final de graduação [1]. Tal trabalho consiste no desenvolvimento de um método de comparação de conteúdos musicais baseado em autômatos adaptativos, servindo de base para um protótipo de um sistema de identificação de melodias. A busca é realizada sobre um repertório previamente estabelecido, a partir de uma amostra de áudio contendo a execução de um pequeno trecho de sua melodia assobiado.

O principal objetivo deste artigo é apresentar o desenvolvimento deste autômato adaptativo para reconhecimento e busca de melodias, assim como alguns resultados preliminares.

Comumente as pessoas se deparam com o problema de tentar identificar uma música a partir do conhecimento de apenas um trecho da mesma, este trecho, no entanto, muitas vezes consiste apenas de uma melodia, ou um ritmo da música original. Esta situação é recorrente em diferentes contextos, como ao tentar identificar uma música ouvida em um filme, um programa de TV, uma rádio, etc.

Este problema em geral é difícil, devido à grande quantidade de músicas existentes e pela falta de ferramentas que possam

auxiliar esta busca. As formas mais comuns de organização de repositórios musicais recaem sobre estilos musicais e nomes, o que não é o suficiente para endereçar o problema apresentado. Além disso, seria extremamente inviável para um indivíduo varrer um extenso repositório musical procurando pela música desejada. A forma mais comum atualmente de se resolver este problema, é com a busca de um especialista, como um vendedor especializado, por exemplo. Assim, percebe-se que as dificuldades apresentadas geram uma grande demanda por técnicas que possam auxiliar esse tipo de busca.

O principal foco do trabalho desenvolvido é a apresentação de uma forma de comparação de melodias baseados em autômatos adaptativos. Para uma avaliação efetiva de tal método de comparação, o protótipo de um sistema mais amplo que se apóia sobre tal técnica foi desenvolvido, a fim de prover um cenário de teste mais concreto.

II. PROXIMIDADE DE MELODIAS

Quando uma pessoa canta uma melodia ou a toca em um instrumento, somos eventualmente capazes de identificar a que música aquela melodia pertence. Nosso cérebro é capaz de reconhecer estas semelhanças mesmo na presença de variações ou imprecisões na melodia que ouvimos.

Um exemplo típico de tais variações é a transposição tonal, em que a melodia é reproduzida com uma variação (logarítmica) fixa na altura de todas as notas, para mais ou para menos. Outro exemplo de variação é a dilatação ou contração das durações das notas que compõe tal melodia.

Em outros casos, há a ocorrência de erros na reprodução, tais como uma nota errada (com altura diferente), ou mesmo a omissão ou adição de notas à melodia original. Estes erros são, em geral, provenientes da incapacidade ou imprecisão do próprio executor, que reproduz a melodia que o mesmo tem em mente.

Tais fatos evidenciam a existência de um conceito subjetivo de proximidade entre melodias, e é exatamente a subjetividade deste conceito que dificulta a implementação da comparação de trechos melódicos por sistemas computacionais.

Assim, no escopo deste trabalho, alguns dos erros possíveis em uma reprodução por um ser humano foram identificados e categorizados, de modo a servir de base para o desenvolvimento da técnica proposta.

Desta forma, os erros abordados são:

1. Transposição tonal
2. Variação temporal

¹ Este artigo é baseado em um projeto final de graduação desenvolvido pelos alunos Pedro R. N. Pedruzzi e Ricardo A. Redder Jr. sob a orientação do Prof. Dr. João José Neto, no Laboratório de Linguagens e Técnicas Adaptativas (LTA EP/USP).

3. Erros simples
 - a. Troca de uma nota
 - b. Adição de uma nota
 - c. Omissão de uma nota

Tais categorias englobam grande parte dos erros cometidos durante a reprodução de uma melodia, e todas elas contribuem (em maior ou menor grau) para o conceito de proximidade de dois trechos de áudio. Percebe-se também, que tais erros só podem ser estabelecidos quando em comparação com uma segunda melodia, que é considerada a correta.

Assim, define-se que a medida de similaridade entre duas melodias será função dos erros (dentre estas categorias) encontrados em uma melodia quando comparada à outra.

Para mensurar os erros em cada categoria diferentes técnicas se fazem necessárias, assim, três técnicas diferentes foram empregadas para endereçar cada problema.

A. TRANSPOSIÇÃO TONAL E VARIAÇÃO TEMPORAL

Em um ambiente sem a presença de erros, pode-se analisar a proximidade entre duas melodias com a mesma quantidade de notas, assumindo que existe uma relação matemática que mapeiam aproximadamente as notas de uma melodia nas notas da outra. Uma relação mapeia as alturas e outra as durações. Sendo p_1 e p_2 , respectivamente as alturas de uma nota do trecho 1 e sua correspondente no trecho 2; e d_1 e d_2 as durações destas; a relação que aproxima o mapeamento de durações é do tipo:

$$p_1 = A.p_2$$

A constante A representa uma proporcionalidade entre as durações, porém permitindo dilatações e contrações.

Para mapear as alturas utiliza-se a seguinte relação:

$$\log d_1 = \log d_2 + B$$

A constante B representa a transposição tonal. A relação logarítmica é necessária pelo fato de que a percepção do ouvido humano para alturas de notas é exponencial.

A partir destas relações de aproximação, calculam-se os parâmetros A e B que melhor aproximam a distribuição, utilizando o método dos mínimos quadrados [2]. O valor dos parâmetros obtidos eventualmente pode ser utilizado para avaliar a proximidade entre as melodias. Porém, nesta modelagem, o relevante não são os parâmetros obtidos da redução, e sim, a soma quadrática dos erros ao utilizá-los:

$$S_p = \sum (A.p_2 - p_1)^2$$

$$S_d = \sum (\log d_2 + B - \log d_1)^2$$

Quanto menor forem estas somas, mais próximas são as melodias comparadas. A distância entre as melodias é definida por uma soma ponderada destas somas, com pesos parametrizáveis.

Maiores detalhes sobre o método de comparação desenvolvido podem ser encontrados em [1].

B. ERROS SIMPLES

Para lidar com erros simples foi empregada uma técnica baseada em autômatos adaptativos [3]. A idéia básica é a construção de um autômato adaptativo a partir de uma das melodias que seja capaz de reconhecer melodias similares a esta, contendo eventuais erros simples devidamente espaçados. Então se avalia a semelhança entre as melodias aplicando o autômato sobre a outra melodia.

O primeiro passo do processo estabelecido consiste em quantizar as alturas das notas das melodias, acomodando-as em uma escala discretizada (que varia de 0 a 127), que corresponde à codificação utilizada em arquivos MIDI. Desta maneira as melodias podem ser vistas como cadeias de símbolos. Note que este processo de quantização não é um simples arredondamento, e envolve a estimativa de um fator de transposição que deve ser aplicado para reduzir ao máximo o erro de quantização. Este método foi chamado de quantização absoluta e foi descrito em [1].

A configuração inicial do autômato concebido é um caminho de estados e transições capaz de reconhecer cadeias de entrada equivalentes àquela que produziu o autômato, a menos de transposição tonal, que é corrigida no início do reconhecimento com base na primeira nota.

Porém ao receber um símbolo neste caminho para o qual não existe transição, ocorre uma ação adaptativa em que uma estrutura de novos estados e transições é incorporada ao autômato a partir do estado corrente. A nova configuração é capaz de contornar as situações de erro previstas (troca, adição ou omissão de uma nota).

Para melhor compreensão do processo considere o seguinte exemplo de comparação entre duas cadeias. Seja a cadeia de base para comparação: 69, 71, 73, 74, 76, 77, 76. A partir desta cadeia, o autômato da Figura 1 é gerado.

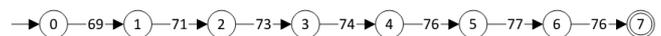


Figura 1 - Configuração inicial do autômato

Dada uma segunda cadeia: 69, 71, 74, 74, 76, 77, 76; que corresponderia à cadeia a ser comparada com a cadeia base. Tal cadeia serve como entrada para o autômato construído. Após consumir os dois primeiros valores (69 e 71) o terceiro valor (74) não pode disparar a transição, e sinaliza um erro. Assim uma ação adaptativa é disparada, gerando uma nova configuração para o autômato, que pode ser visto na Figura 2. Após esta ação, o reconhecimento continua, neste caso, o quarto valor seria consumido, disparando a transição do estado 8 para o estado 4, e em seguida o restante da cadeia seria consumido normalmente atingindo o estado.

Cada uma das três novas transições geradas ($8 \rightarrow 3$, $8 \rightarrow 4$ e $8 \rightarrow 5$) indicam um tipo de erro: adição, troca ou omissão, respectivamente. Desta forma, o fato da cadeia seguir o caminho $8 \rightarrow 4$ indica que houve a troca de uma nota, ou seja, a troca da nota 73 (terceira nota) por uma nota 74.

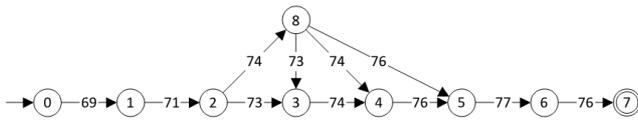


Figura 2 - Configuração após adaptação

Desta forma pode-se dizer que a distância entre as duas melodias em questão é de um erro simples, mais especificamente uma troca.

III. SISTEMA DE BUSCA

Para avaliar os métodos de comparação propostos o protótipo de um sistema de busca foi desenvolvido, de forma a englobar as partes mais relevantes de um sistema concreto.

O protótipo apresentado se limita a receber um sinal de áudio (um assobio) de um usuário, converter tal sinal para uma sequência de notas, e em seguida comparar tal sequência com um repositório de músicas previamente estabelecido. Através das comparações efetuadas pode-se estabelecer uma medida de distância entre cada música do repositório e o sinal de entrada, e assim gerar uma lista das músicas que mais se assemelham ao sinal de entrada.

O fluxo de informações do sistema pode ser visto na Figura 3.

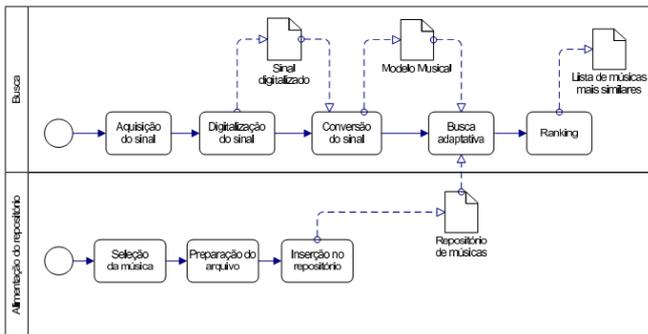


Figura 3 - Fluxo de informações do sistema

Neste sistema, inicialmente um sinal gerado pelo usuário (um assobio) é gravado, correspondendo a sua interpretação de uma música. Este sinal é digitalizado, e em seguida serve de entrada para um componente de conversão que converte o sinal digitalizado, contendo as frequências do sinal original, em um modelo musical que representa as notas assobiadas (início, duração e frequência de cada nota). O modelo musical gerado é quantizado de acordo com o método apresentado anteriormente. A partir da cadeia de notas quantizadas um processo de busca sobre as músicas do repositório é efetuado, utilizando-se o método de comparação adaptativa apresentado. Após esta comparação pode-se apresentar como resposta do sistema uma lista contendo as músicas do repositório mais similares à melodia de entrada, ordenadas pela similaridade.

A. BUSCA

Com base na definição de distância entre melodias apresentada na seção anterior é possível construir um mecanismo de busca simples de uma melodia sobre um repositório de músicas, comparando-se a melodia de entrada com todas as músicas do repositório. O processo segue da seguinte forma:

A partir do modelo musical gerado pelo componente de conversão, e após sua quantização, um autômato análogo ao da Figura 1 é gerado, esta é a cadeia que deve ser comparada com as músicas do repositório.

A comparação é feita aplicando-se ao autômato todas as melodias do repositório de maneira a utilizar uma janela deslizante; ou seja, para uma dada música, o autômato recebe inicialmente a primeira nota da cadeia representando a música sendo analisada, e consome a cadeia até que não seja mais possível continuar (caso encontre mais de um erro simples seguido), ou chegue ao final do autômato. Em seguida, o mesmo processo é repetido, porém tomando como base a nota seguinte, no caso, a segunda nota.

O caminho mais curto entre o estado inicial e o final é obtido quando a cadeia de entrada equivale exatamente à sequência de notas da cadeia procurada. Porém, outras transições existem para flexibilizar a comparação, contornando alguns tipos de erros. Por este motivo, o autômato é capaz de reconhecer outras cadeias que não correspondem à cadeia exata da procurada.

Durante a execução deste processo, os tipos de erros identificados e suas localizações na cadeia são registrados. Desta forma, para uma dada música, o menor valor obtido (menor número de erros) é dado como o valor de erro para a música em questão.

Após a execução deste processo para todas as músicas do repositório é possível estabelecer para cada música o trecho mais próximo da melodia de entrada, além da quantidade de erros em relação a esta.

Com posse destas informações é possível reverter os erros gerados, e assim produzir uma melodia artificial, baseada na melodia de entrada, porém desfazendo-se os erros percebidos. Desta forma, se uma nota foi omitida, por exemplo, pode-se adicioná-la à melodia novamente. Isso possibilita a realização da comparação no domínio temporal apresentada anteriormente, obtendo-se assim mais uma medida de erro que pode auxiliar no estabelecimento da similaridade entre as melodias.

Ao fim da comparação da melodia de entrada com todas as músicas do repositório, têm-se algumas medidas que podem auxiliar no estabelecimento da similaridade entre a melodia de entrada, e cada música do repositório:

1. Número de erros entre a melodia e a música
Proveniente da execução do autômato.
2. Aceitação ou não do autômato
Proveniente da execução do autômato.
3. Valores numéricos de soma de distâncias
Proveniente da comparação numérica.

B. ORDENAÇÃO

Os valores de cada medida apresentadas anteriormente influenciam na similaridade entre a melodia de entrada e uma música dada. Porém a importância que deve ser dada a cada dimensão não é clara, assim não é possível estabelecer de imediato uma fórmula que sumarie todos os dados obtidos.

Desta forma, a fim de avaliar o desempenho do método de comparação adaptativa, apenas os dados provenientes do autômato foram utilizados para ordenação.

IV. RESULTADOS

Inicialmente foi criado um repositório contendo 37 músicas, de diversos gêneros e todas em formato MIDI. Em seguida, 19 amostras de áudio (assobios) foram colhidas, todas de músicas existentes no repositório.

Todas as amostras foram submetidas ao sistema construído, colhendo-se a resposta dada. Conforme explicado na seção anterior, contabilizaram-se apenas os resultados provenientes do autômato, desta forma procurava-se ter uma idéia da eficiência de comparação do autômato proposto.

Por fim, as respostas foram analisadas, procurando-se dentro de cada resposta, a posição que a melodia procurada realmente encontrava-se. Ou seja, para um assobio correspondendo a uma música A, procurou-se na resposta do sistema qual a posição da música A na lista ordenada gerada. A compilação desta análise pode ser vista no gráfico da Figura 4.

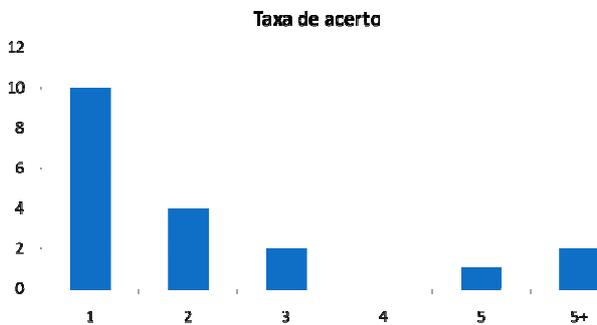


Figura 4 - Taxa de acerto para o conjunto de testes

O gráfico mostra que na maior parte dos casos o sistema identificou corretamente a música procurada como a mais similar à melodia de entrada.

Em outros casos, apesar da música desejada não corresponder a primeira colocação, ainda assim encontrava-se entre as três primeiras colocações.

V. TRABALHOS FUTUROS

Após a conclusão desta fase de trabalhos, alguns avanços e melhorias necessários se fazem evidentes:

1. Maior granularidade de erros
Apenas três tipos de erros foram considerados para a criação do autômato proposto (omissão, adição e troca), porém percebeu-se que ao menos uma nova

classe de erros se faz necessária, uma classe que indique uma pequena variação com relação à nota em questão, dando uma granularidade maior ao tipo de erro de troca.

2. Múltiplos erros
Apenas erros simples foram considerados, que são os erros apresentados anteriormente, casos de erros seguidos não foram abordados, e levam ao fim da execução do autômato. Erros mais complexos poderiam ser considerados, por exemplo: TROCA, TROCA, OMISSÃO. A metodologia proposta não seria capaz de lidar com tais variações.
3. Definição de pesos para cada classe de erro
Conforme citado anteriormente, diversos fatores contribuem para a variação da distância entre dois trechos de áudio, porém, não é claro a forma como dada um destes fatores contribui, assim, mais estudos, e avaliações se fazem necessárias para a determinação dos pesos mais adequados para cada dimensão.
4. Generalização do método de comparação
O método de comparação proposto independe do fato da cadeia ser proveniente de uma sequência de notas, o que abre a possibilidade de generalização deste método para lidar com cadeias de símbolos de qualquer natureza. Uma generalização de todos os conceitos envolvidos se faria necessária, como alfabeto de entrada, classes de erros, etc. Tal método de comparação seria adequado para buscas de padrões em cadeias.

VI. CONCLUSÕES

Ao final do desenvolvimento deste trabalho pode-se perceber que a técnica proposta é consistente e é capaz de lidar de maneira eficiente com distúrbios comuns em reproduções de áudio por seres humanos, como trocas de nota ou omissões.

Percebe-se também que a combinação de diversas técnicas, a fim de mensurar diferentes dimensões de variações nos trechos comparados, parece ser o caminho mais indicado.

Além disso, fica evidente que o método ainda é prematuro para uso em sistemas de larga escala, porém a abordagem diferenciada do problema, ou seja, o uso de uma técnica adaptativa sobre tal campo, abre interessantes possibilidades.

REFERÊNCIAS

- [1] Pedruzzi, P. R. N., e Redder, R. A. "Reconhecimento e Busca Adaptativos de Padrões Musicais". Thesis. Escola Politécnica da Universidade de São Paulo, 2008.
- [2] Otto Bretscher (1995). "Linear Algebra With Applications", 3rd ed.. Upper Saddle River NJ: Prentice Hall.
- [3] J. J. Neto, "Adaptive Automata for Context-Sensitive Languages", SIGPLAN NOTICES, Vol. 29, n. 9, September, 1994, pp. 115-124.