IDENTIFICAÇÃO DE NOTAS MUSICAIS EM REGISTROS SOLO DE VIOLÃO E PIANO

Alexandre Leizor Szczupak

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Aprovada por:

_____
Prof. Luiz Wagner Pereira Biscainho, D.Sc.

_____
Prof. Luiz Pereira Calôba, Dr.Ing.

_____
Prof. Sérgio Lima Netto, Ph.D.

_____
Prof. Marcio Nogueira de Souza, D.Sc.

RIO DE JANEIRO, RJ - BRASIL
JUNHO DE 2008

SZCZUPAK, ALEXANDRE LEIZOR

Identificação de Notas Musicais em Registros Solo de Violão e Piano

[Rio de Janeiro] 2008

IX, 123 p., 29,7 cm (COPPE/UFRJ, M.Sc., Engenharia Elétrica, 2008)

Dissertação - Universidade Federal do Rio de Janeiro, COPPE

1.Transcrição Musical Automática 2.Redes Neurais 3. Transformada de Q Constante

I. COPPE/UFRJ II. Título (série)

Agradecimentos

Agradeço aos meus orientadores, Luiz Wagner Pereira Biscainho e Luiz Pereira Calôba e aos amigos, Tadeu Nagashima Ferreira, Alan Freihof Tygel, Filipe Castello da Costa Beltrão Diniz, Fábio Pacheco Freeland, Leonardo de Oliveira Nunes, Rafael Almeida de Jesus, Flávio Rainho Ávila, Rafael Cauduro Dias de Paiva, Rafael Andrade Santos Pantoja, Jorge Costa Pires Filho, Iúri Kothe, Lisandro Lovisolo, Michel Pompeu Tcheou, Alessandro J. Salvaterra Dutra, Ana Luisa A. Santos, Amaro Azevedo de Lima, Wallace Alves Martins, Markus Vinícius Santos Lima, Gustavo Luis Almeida de Carvalho, Rodrigo C. Meirelles, Jose Fernando Leite de Oliveira, Arnaldo Satoru Gunzi, Maurício de Carvalho Machado, Daniele Cristina Oliveira da Silva, Paulo Antônio Andrade Esquef, Lara Christiana R. L. Feio, Michelle de Araújo Nogueira, Luciana Requião, Pedro Lucio Bittencourt e Renato Baran.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

IDENTIFICAÇÃO DE NOTAS MUSICAIS EM REGISTROS SOLO DE VIOLÃO E PIANO

Alexandre Leizor Szczupak

Junho/2008

Orientadores: Luiz Wagner Pereira Biscainho
Luiz Pereira Calôba

Programa: Engenharia Elétrica

Nesta dissertação são apresentados métodos desenvolvidos para a identificação de notas musicais em registros de violão solo. Estes métodos têm como base o uso de redes neurais *feed-forward* de múltiplas camadas, treinadas com representações espectrais obtidas através de uma transformada de Q constante. Além destes, também são apresentadas adaptações voltadas para a identificação de notas musicais em registros de piano.

Os métodos podem ser divididos em duas abordagens: na primeira, apenas uma rede é utilizada na identificação das notas presentes em cada segmento de sinal analisado; na segunda, duas redes são utilizadas em seqüência: a primeira para identificar apenas a nota mais grave de cada segmento de sinal analisado e a segunda para encontrar os intervalos entre a nota mais grave e as notas restantes.

Os resultados dos métodos desenvolvidos para violão foram promissores, porém, os resultados das adaptações para piano não foram bons. Para ambos os casos, os melhores resultados foram obtidos através da segunda abordagem, principalmente no desempenho isolado da etapa de identificação de intervalos entre a nota mais grave de cada segmento de sinal e as notas restantes.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

MUSICAL NOTES IDENTIFICATION IN SOLO RECORDINGS OF ACOUSTIC GUITAR AND PIANO

Alexandre Leizor Szczupak

June/2008

Advisors: Luiz Wagner Pereira Biscainho
Luiz Pereira Calôba

Department: Electrical Engineering

This dissertation presents methods developed for the identification of musical notes in acoustic guitar recordings. These methods are based on multilayer feed-forward neural networks, trained with frequency domain representations obtained via a constant-Q transform. Versions of these methods, developed for the identification of musical notes in piano recordings, are also presented.

The proposed methods can be divided in two categories: methods based on a single neural network, used to identify the notes in a signal excerpt; and methods with two neural networks used in sequence, the first one to identify the bottom note of a signal excerpt and the second to determine the intervals between the bottom note and the remaining ones.

Encouraging results were obtained on the identification of musical notes in acoustic guitar recordings, but not on the identification of musical notes in piano recordings. For both instruments, the best results were obtained using methods of the second category, especially regarding the isolated performance of the neural network used to determine intervals between the bottom note and the remaining ones.

Sumário

Capítulo 1

Introdução

1.1 Transcrição Musical Automática

Transcrição musical é um processo de atribuição de símbolos para eventos selecionados de um sinal musical. A atribuição deve ser realizada de modo que torne possível reproduzir a qualidade e a seqüência destes eventos. Este processo pode ser comparado com uma codificação de sinal, com interesse não na recuperação do sinal original, mas na recriação dos eventos com um novo sistema (instrumento musical) para obter um novo sinal que caracterize, de acordo com a percepção humana, a música transcrita.

Diversos tipos de eventos podem ser registrados, e a escolha de quais devem constar na transcrição depende da forma que ela será apresentada. Formas comuns para a transcrição são: partituras musicais, cifras, tablaturas e arquivos MIDI. Cada uma delas possui um conjunto diferente de símbolos para representar os eventos de uma música. Alguns eventos importantes são: as notas tocadas, os instantes em que se iniciam (*onsets*), suas durações e os intervalos (diferenças de altura) entre notas simultâneas. Também pode ser importante representar informações auxiliares como o andamento da música, a escala musical e o compasso.

Nesta tese o foco é a identificação, ao longo do tempo, de notas musicais. O objetivo é obter um método computacional para a identificação das notas presentes em gravações solo de violão e piano. Também deve ser possível adaptar o método para uso com gravações de outros instrumentos musicais polifônicos com *pitch* determinado.

1.2 Polifonia

No contexto da análise de gravações de instrumentos musicais solo, a polifonia é entendida como a presença de mais que uma nota simultaneamente em um trecho de sinal.[1] O grau de polifonia de um violão é igual ao número de cordas do instrumento. Um músico tocando um violão de 6 cordas pode produzir até 6 notas simultaneamente. No caso de um piano, na execução tradicional, um músico pode acionar até 10 teclas simultaneamente, porém mais notas podem soar ao mesmo tempo. Como cada nota do piano tem um mecanismo independente, tocar uma nota não interrompe o som de outras que já estejam soando. Assim, o grau de polifonia do piano é, no limite, igual ao número de suas teclas.

1.3 *Pitch*

O termo *pitch* usualmente se refere à freqüência da onda senoidal que é melhor associada, perceptivamente, a um dado som. Na análise de sinais de música, *pitches* podem ser associados aos sons da voz cantada e aos sons de muitos instrumentos musicais. De acordo com HERRERA-BOYER *et al.* [1], instrumentos com sons que provocam sensação evidente de *pitch* (como os cordofones e os aerofones)[2] são chamados de "*pitched*", "afinados" ou "com *pitch* determinado". Instrumentos que não têm *pitch* definido (como a maioria dos idiofones e dos membranofones)[3] são chamados de "*unpitched*", "sem afinação" ou "com *pitch* indeterminado".

Alguns instrumentos fortemente inarmônicos, como os pratos de bateria, dificilmente podem ser usados para gerar sons com *pitch* evidente, e por isto são classificados como instrumentos de *pitch* indeterminado. Porém, outros que recebem esta mesma classificação podem provocar sensações evidentes de *pitch*, embora não

[1] Para a teoria musical, a polifonia é definida como a combinação de duas ou mais linhas melódicas.

[2] Cordofones são instrumentos com atuação sobre cordas, como violões e pianos. Aerofones são instrumentos com atuação sobre colunas de ar, como flautas e trompetes.

[3] Idiofones são instrumentos com atuação sobre o próprio corpo do instrumento, como pratos de bateria e marimbas. Membranofones são instrumentos com atuação sobre uma membrana elástica, como surdos e atabaques.

sejam projetados para gerar sons de acordo com escalas de valores pré-estabelecidos de *pitch*. Um exemplo é o surdo de bateria, que pode ser afinado para evidenciar um *pitch* de uma escala pré-determinada, apesar de isto normalmente não ser feito. Sua afinação em geral é feita buscando apenas estabelecer razões de *pitch* entre seu som e os de outros membranofones tocados em conjunto com ele, que também têm *pitch* indeterminado. Na afinação da maioria dos membranofones, as razões entre *pitches* (intervalos musicais) não seguem regras pré-estabelecidas, ficando ao gosto do músico. Já no caso dos instrumentos com *pitch* determinado, as razões entre *pitches* de diferentes notas seguem regras de acordo com o projeto do instrumento e de acordo com o temperamento da escala de *pitches* escolhida.

Apesar de alguns autores utilizarem os termos *pitch* e freqüência fundamental ($f0$) indistintamente, estabelecer a diferença entre eles é importante para a análise de sinais musicais, particularmente na análise de sinais de cordofones, sujeitos a fortes efeitos de inarmonicidade.

1.4 Temperamento

Na teoria musical não se utilizam diretamente valores de *pitch* para designar a altura dos sons. Em vez disto, cada altura é indicada como uma nota musical e cada nota tem um valor associado a ela. A escala é definida a partir da determinação de um *pitch* para uma nota de referência e do uso de uma regra de temperamento.

Alguns exemplos de regras de temperamento são: a afinação justa, a afinação pitagórica e a afinação em temperamento igual. Todas têm em comum a subdivisão de oitavas em 12 intervalos, chamados semitons, porém cada regra determina de forma diferente a extensão de cada intervalo da escala. A regra mais comum na música ocidental contemporânea é a de temperamento igual. Nela, os *pitches* das notas são dispostos em uma progressão geométrica com razão $q = 2^{1/12}$. Para obter o *pitch* de uma nota n semitons mais alta ou mais baixa que a nota de referência, deve-se, respectivamente, multiplicar ou dividir o *pitch* da nota de referência por q^n. Desta forma, cada oitava abrange 12 notas com *pitches* igualmente espaçados em escala logarítmica. Uma referência comumente utilizada é a nota Lá 4 [4], com *pitch*

[4]Nesta tese, adota-se como convenção nomear a primeira oitava dos pianos comuns, de 88 teclas,

igual a 440 Hz.

Os *pitches* de notas geradas com instrumentos musicais reais normalmente não recaem perfeitamente sobre os valores definidos no temperamento utilizado. Na prática recaem sobre pontos na vizinhança destas freqüências. Alguns instrumentos, inclusive, possibilitam utilizar como recurso estético uma modulação cíclica do *pitch*, o *vibrato*.

1.5 Inarmonicidade em Pianos e Violões

Um dos problemas presentes na identificação de notas musicais, a inarmonicidade ocorre quando um som não tem suas parciais[5] ordenadas em uma série harmônica, isto é, suas parciais não são ordenadas em uma progressão aritmética com razão igual à freqüência fundamental. Para cordofones ela é caracterizada por desvios positivos nas freqüências das parciais em relação às freqüências harmônicas. Estes desvios se devem à rigidez elástica do material. As freqüências das parciais de uma corda real sem enrolamento[6] podem ser obtidas, em função de sua freqüência fundamental f_0, através das equações abaixo [2]:

$$f_n = n f_0 \sqrt{1 + Bn^2} \tag{1.1}$$

$$B = \frac{\pi^3 E d^4}{64 l^2 T}, \tag{1.2}$$

onde as propriedades da corda são:

$B =$ coeficiente de inarmonicidade,

$E =$ módulo de Young,

$d =$ diâmetro,

$l =$ comprimento e

$T =$ tensão da corda.

---

de 'oitava 0'. A primeira nota do piano, uma nota Lá, é então chamada de Lá 0. A nota Lá 4 é a 49^a nota nos pianos de 88 teclas e fica na oitava 4.

[5] Parcial aqui se refere a cada uma das componentes senoidais que modelam o sinal.

[6] Como artifício para abaixar o *pitch* de uma corda sem aumentar excessivamente seu diâmetro ou comprimento, ela pode ser fabricada envolta por um enrolamento metálico.

O coeficiente de inarmonicidade B assume valores não-negativos. Quando $B = 0$, a relação entre as parciais é perfeitamente harmônica, mas para um B positivo, o desvio das parciais cresce com n. As equações (1.1) e (1.2) são de fato aproximações, válidas apenas se o deslocamento transversal da corda estiver restrito a uma pequena região em torno da posição de equilíbrio. Um modelo completo precisaria levar em conta que a tensão sobre a corda, bem como seu comprimento, variam não-linearmente com seu deslocamento transversal, e que seus modos de vibração também dependem da rigidez dos suportes em suas extremidades [3]. Quanto maior for a rigidez do suporte, menor sua influência no deslocamento das parciais.

Os efeitos da variação do coeficiente de inarmonicidade B na percepção do *pitch* foram estudados por JÄRVELÄINEN *et al.* [2]. Os autores realizaram testes perceptivos utilizando sons de 4 notas de piano, sintetizados através de modelagem senoidal. Nos testes, freqüências parciais de cada nota sintetizada foram ajustadas seguindo variações nos valores de B, de acordo com a Equação (1.1). Para valores de B próximos a zero, os *pitches* se mantiveram próximos aos valores das freqüências fundamentais, porém cada nota também apresentou uma faixa particular de valores na qual o aumento de B foi acompanhado pelo aumento do *pitch*. Valores do coeficiente de inarmonicidade além destas faixas geraram resultados ambíguos: alguns indivíduos igualaram os *pitches* novamente a valores próximos das freqüências fundamentais, enquanto outros continuaram acompanhando os incrementos na inarmonicidade.

Os conjuntos de parciais de duas versões de uma mesma nota, tocadas da mesma forma em dois instrumentos iguais exceto por cordas de modelos distintos, podem ser diferentes por causa das mudanças nos valores de E.

A distância entre os suportes das cordas de um violão, a ponte (localizada no tampo superior do instrumento) e o capotraste (localizado no braço do instrumento), depende do modelo do instrumento. Quanto maior é esta distância, menor é a inarmonicidade da corda, como indicado na Equação (1.2).

Cada nota de piano é gerada pela vibração simultânea de uma, duas ou três cordas de mesmo comprimento. Os valores de E das cordas associadas a uma determinada nota, podem variar não só de modelo para modelo do instrumento, como entre as próprias cordas. Os comprimentos das cordas também variam muito entre

diferentes modelos do instrumento. Cordas mais longas são utilizadas para obter menor inarmonicidade, fator importante na caracterização do timbre do instrumento.

1.6 Não-Linearidades

O comprimento de uma corda vibrante, presa entre dois suportes fixos, varia não-linearmente acompanhando seu movimento oscilatório. O efeito desta não-linearidade sobre as parciais se torna relevante quando a amplitude de vibração é grande, como acontece quando um músico usa dinâmica *forte* ou *fortissimo*. Nesta situação, o deslocamento transversal inicial da corda, que precede sua oscilação, é muito grande. Isto aumenta transitoriamente a tensão média da corda. Por isto, as freqüências de todos os modos iniciam em valores ligeiramente superiores aos previstos na Equação (1.1) e, acompanhando a variação na amplitude das oscilações, decaem para os valores previstos para oscilações de baixa amplitude [4].

LEGGE e FLETCHER [3] demonstraram a presença de não-linearidades de segunda e terceira ordem nos sistemas formados por cordas vibrantes montadas em suportes não perfeitamente rígidos[7]. Entre os efeitos destas não-linearidades estão o surgimento, ao longo do tempo, de modos de vibração inicialmente não excitados nas cordas e de acoplamentos entre os modos presentes. Após o *onset* de uma nota, os modos de vibração da corda que têm um nó sobre a posição do golpe têm amplitudes iniciais próximas a zero. Devido aos acoplamentos, as amplitudes destes modos crescem e atingem valores máximos após um período de tempo da ordem de 0,1 s. Os acoplamentos, associados à inarmonicidade das cordas, também causam flutuações nas amplitudes de todos os modos de vibração presentes.

As freqüências das parciais presentes na saída de um sistema não-linear são dadas, para cada termo de ordem η que modela o sistema, por todas as somas possíveis, tomadas η a η, entre as freqüências das parciais presentes na entrada [5]. Assim, o número de parciais observadas na saída de um sistema não-linear é maior que o número de parciais observadas na entrada do sistema. Isto é comum nos

[7]Um suporte perfeitamente rígido tem admitância mecânica igual a zero. A admitância mecânica Y é definida como $Y(\omega) = V(\omega)/F(\omega)$, sendo ω = freqüência, V = velocidade e F = força.

períodos de ataque e decaimento das notas[8], que concentram a maior parte dos efeitos de não-linearidades sobre o sinal.

Numa análise espectral feita com alta resolução é possível observar que cada modo das cordas é na verdade um par de modos com freqüências muito próximas [7]. A vibração de cada corda pode ser decomposta em dois planos de polarização. Como os suportes das cordas têm impedâncias acústicas diferentes em cada plano, os comprimentos efetivos das cordas (e as freqüências dos modos de vibração) diferem ligeiramente para cada plano.

1.7 O Violão

Um violão comum de seis cordas tem extensão de 44 notas, de Mi 2 até Si 5. Numa execução tradicional, as notas podem soar individualmente ou em combinações de duas até seis notas simultâneas. O posicionamento de trastes ao longo do braço serve para estabelecer os nós de vibração necessários para gerar cada nota. Das 44 notas, 34 podem ser obtidas a partir de pelo menos duas posições distintas do braço. As cinco notas mais graves e as cinco mais agudas do instrumento só podem ser obtidas, cada uma, a partir de uma única posição sobre o braço.

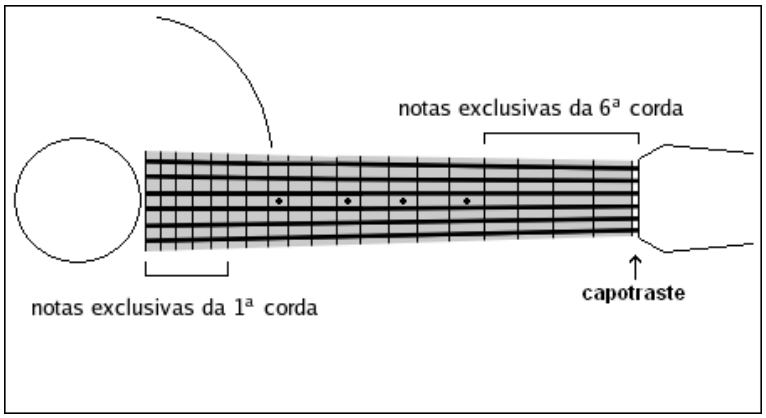

Figura 1.1: Representação do braço de um violão. As cordas mais agudas ficam nas posições inferiores.

O braço do violão é projetado para manter em progressão geométrica as

[8]A envoltória de uma nota musical é comumente descrita através de um modelo ADSR [6], formado por uma seqüência de quatro períodos: Ataque, Decaimento, Sustentação e Liberação (*Attack, Decay, Sustain and Release*)

freqüências fundamentais das notas adjacentes de uma mesma corda. Esta progressão, como a das escalas igualmente temperadas, tem razão $q = 2^{1/12}$. Assim, se a freqüência fundamental de vibração da 5^a corda solta (Lá 2) é igual a 110,00 Hz, a freqüência fundamental da nota adjacente mais alta (Lá♯2, um semitom acima) é $110{,}00 \text{ Hz} \times 2^{1/12} \approx 116{,}54 \text{ Hz}$. Pode-se notar que a progressão geométrica entre as freqüências fundamentais pode ser obtida pelo correto posicionamento dos trastes, porém a razão entre os *pitches* de notas adjacentes de uma mesma corda depende do valor do módulo de Young (E) da corda. Quanto menor for o valor de E, mais próximo de q será a razão entre *pitches* adjacentes de uma mesma corda e maior será a correlação entre estes *pitches* e a escala de temperamento igual. Se E for muito elevado, grandes desvios de *pitch* podem ocorrer em comparação com as freqüências previstas nas escalas de temperamento igual, principalmente nas notas obtidas utilizando pequenos segmentos de corda.

Cada modelo de violão apresenta um conjunto particular de ressonâncias, de acordo com seu projeto. As duas ressonâncias mais influentes na sonoridade da maioria dos modelos são a A0 (ressonância de Helmholtz)[9] e a T1 (1^{a} ressonância do tampo superior).

A ressonância A0 resulta da interação do sistema formado pelo corpo do instrumento com a cavidade do tampo superior, já a ressonância T1 se dá no modo fundamental de vibração do tampo superior [8, 9]. Para medir A0, a vibração do tampo superior deve ser impedida, enquanto para medir T1, a cavidade do tampo superior deve ser fechada. Desta forma evita-se o acoplamento dos dois sistemas. Em um violão livre (sem restrições de vibração e com a cavidade desimpedida), as duas primeiras freqüências de ressonância resultam da interação entre os sistemas e diferem das freqüências de ressonância A0 e T1 [8].

A 1^{a} freqüência de ressonância de um violão livre é localizada tipicamente dentro da faixa entre 70 Hz e 140 Hz. A posição exata depende do modelo do instrumento. Esta faixa sobrepõe parcialmente a faixa de freqüências fundamentais da 1^{a} oitava do violão. A 2^{a} ressonância, geralmente de menor intensidade, tem

[9] O símbolo da ressonância de Helmholtz, A0, também é utilizado para representar, de acordo com o padrão norte-americano de notação musical, a nota musical Lá 0, mas não há relação entre os dois conceitos.

freqüência próxima ao dobro da freqüência da 1ª ressonância. Contudo, enquanto a 1ª freqüência de ressonância do violão livre pode ter um desvio significativo em relação à freqüência de A0, a freqüência da 2ª ressonância do violão livre se mantém próxima do valor da freqüência de T1.

Parciais causadas por estas ressonâncias podem ser observadas no espectro dos sinais de violão, principalmente no período inicial de ataque e decaimento das notas, onde uma grande quantidade de modos do instrumento são excitados graças à natureza impulsiva do golpe sobre a corda.

A Figura 1.2 contém a forma de onda dos 0,30 s iniciais de um registro de violão (nota Dó♯3). Nas Figuras 1.3 e 1.4 são mostrados os valores absolutos de DFTs (*discrete Fourier transforms*) dos trechos entre 0,0 s e 0,15 s e entre 0,15 s e 0,30 s deste registro.

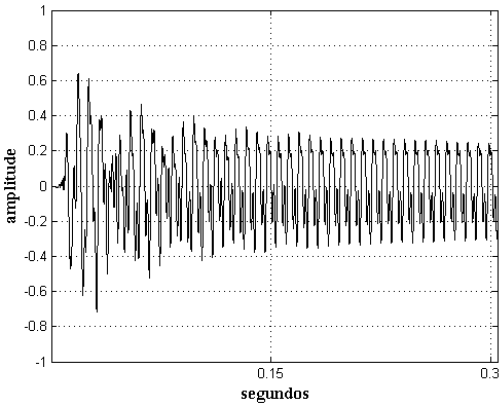

Figura 1.2: Forma de onda dos 0,30 s iniciais de um registro da nota Dó♯3.

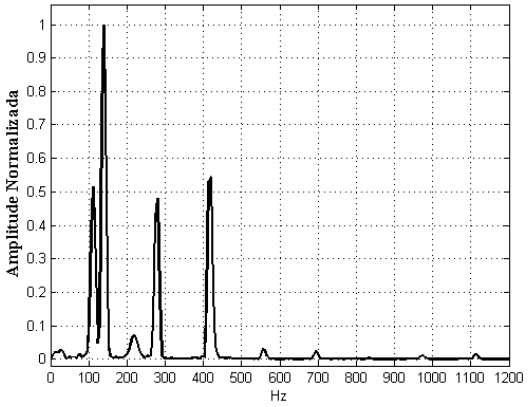

Figura 1.3: DFT do trecho entre 0,0 s e 0,15 s do registro da nota Dó♯3.

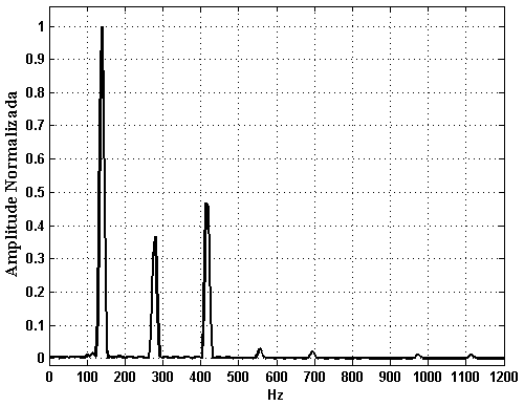

Figura 1.4: DFT do trecho entre 0,15 s e 0,30 s do registro da nota Dó♯3.

Na DFT do primeiro trecho (Figura 1.3) é possível observar não só as parciais da série harmônica de Dó♯3 ($f0 \approx 138{,}59$ Hz) como também duas outras parciais em aproximadamente 110 Hz e 220 Hz. A primeira ocorre na 1ª freqüência de ressonância do violão, e a segunda, de menor amplitude, na 2ª freqüência de ressonância do

instrumento. A segunda DFT (Figura 1.4), realizada sobre um trecho do período de sustentação da nota, não apresenta mais as parciais devidas às freqüências de ressonância do violão.

A análise de trechos de sinal contidos em períodos de ataque ou decaimento de notas de violão pode revelar a existência desta formação de parciais, em pares e em razão harmônica, mas que não pertencem às notas buscadas. No caso acima as freqüências das parciais causadas pelas ressonâncias do violão são próximas às freqüências das duas primeiras parciais da nota Lá 2, porém esta nota não faz parte do sinal.

As parciais devidas às ressonâncias do violão podem variar de amplitude de acordo com a intensidade e com quais notas são tocadas, mas suas freqüências são fixas.

A Figura 1.5 contém a forma de onda dos 0,30 s iniciais de outro registro do mesmo violão, desta vez da nota Sol♯3. Nas Figuras 1.6 e 1.7 são mostrados os valores absolutos de DFTs dos trechos entre 0,0 s e 0,15 s e entre 0,15 s e 0,30 s deste registro.

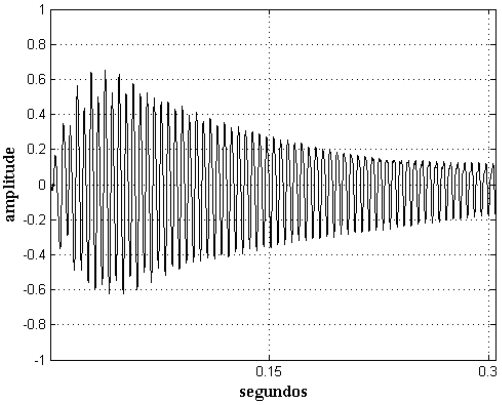

Figura 1.5: Forma de onda dos 0,30 s iniciais de um registro da nota Sol♯3.

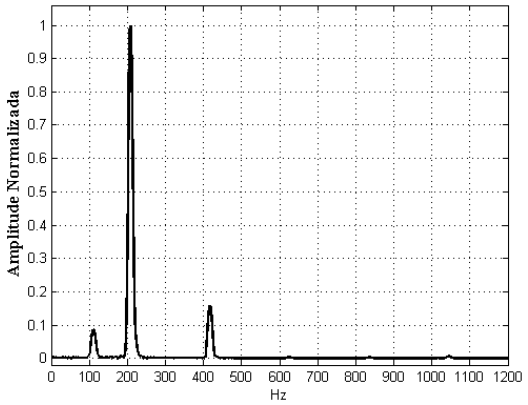

Figura 1.6: DFT do trecho entre 0,0 s e 0,15 s do registro da nota Sol♯3.

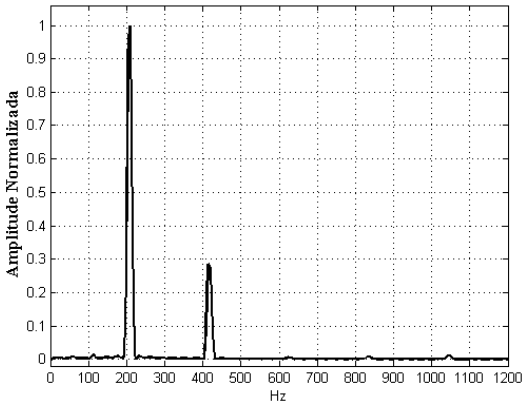

Figura 1.7: DFT do trecho entre 0,15 s e 0,30 s do registro da nota Sol♯3.

Na DFT do primeiro trecho (Figura 1.6) é novamente possível observar uma

parcial em aproximadamente 110 Hz, causada pela 1^a ressonância do violão. Nesta figura, não é possível distinguir a fundamental da série de Sol♯3 ($f0 \approx 207{,}65$ Hz) da parcial da 2^a ressonância do instrumento, em aproximadamente 220 Hz. A segunda DFT (Figura 1.7), realizada sobre um trecho do período de sustentação da nota, novamente não apresenta parciais devidas às freqüências de ressonância do instrumento.

1.8 O Piano

Um piano comum tem extensão de 88 notas, de Lá 0 até Dó 8. Cada nota do instrumento é gerada através do golpe de um martelo sobre um conjunto diferente de cordas. Quando uma tecla do instrumento é pressionada, aciona o martelo correspondente através de um intrincado mecanismo que garante ao músico o controle da dinâmica da nota. O martelo atinge simultaneamente todas as cordas associadas à tecla, afinadas em uníssono.

A quantidade de cordas utilizadas no mecanismo de geração de cada nota depende do projeto do instrumento. De forma geral, são utilizadas de uma a três cordas por nota. Os mecanismos das notas mais graves utilizam menos cordas que os mecanismos de notas mais agudas.

Durante a afinação de um piano, deve ser feita a compensação do efeito de elevação do *pitch* causado pela presença de inarmonicidade nas cordas [2]. A afinação é iniciada a partir de uma nota da região central, como o Lá 4 ($f0 = 440$ Hz), em geral utilizando um diapasão. Em seguida, uma seqüência de intervalos sobre a região central do piano é ajustada, buscando-se estabelecer batimentos pré-definidos entre as parciais de determinadas notas [10]. Por exemplo: a segunda parcial da nota Dó 4 deve, em combinação com o terceiro harmônico da nota Fá 3, gerar um batimento audível de aproximadamente 0,59 Hz. Após todas as notas da região central estarem afinadas, é realizada a afinação das regiões mais graves e mais agudas do instrumento. Para afinar uma nota uma oitava acima de outra nota já afinada, é forçada a coincidência de sua primeira parcial com a segunda parcial da nota já afinada. Para afinar uma nota na oitava abaixo de uma nota já afinada, é forçada a coincidência de sua segunda parcial com a primeira parcial da nota já afinada. Este

procedimento é repetido para todas as notas restantes.

Assim, diz-se que o piano, afinado pela percepção auditiva, tem "escala alongada" [11], caracterizada principalmente por uma distorção na seqüência de frequências fundamentais de suas notas em relação às freqüências da escala de temperamento igual. As freqüências fundamentais de notas próximas à região central do piano acompanham aproximadamente as freqüências da escala de temperamento igual; porém, conforme se observam notas cada vez mais graves, o decréscimo nas freqüências fundamentais se torna mais rápido que o decréscimo dos valores da escala de temperamento igual. No outro sentido, conforme se observam notas cada vez mais agudas, o incremento nos valores das freqüências fundamentais se torna mais rápido que o incremento nos valores da escala de temperamento igual.

Na análise do espectro de uma nota de piano pode ser observada, além da série de parciais com freqüências regidas pela Equação (1.1), a presença de um segundo grupo de parciais. Existem diferentes teorias sobre como deve ser realizada a modelagem das parciais deste grupo [12, 13]. Para CONKLIN [14, 15], que as chama de *phantom partials*, elas ocorrem em dois subgrupos: um formado em freqüências iguais ao dobro das freqüências regidas pela Equação (1.1) e outro em freqüências dadas pelas somas e diferenças de freqüências da mesma equação, tomadas duas a duas. Segundo NAKAMURA e NAGANUMA [16], as freqüências deste grupo de parciais podem ser aproximadas por uma série como a da Equação (1.1), porém com um coeficiente de inarmonicidade igual a $B/4$. WOODHOUSE [7] reportou a presença em registros de violão de parciais que não se enquadram na Equação (1.1) e sugeriu que elas também podem ser modeladas como *phantom partials*.

Outra característica do espectro de notas de piano é a possível presença de parciais causadas por modos de vibração longitudinal. O primeiro destes modos, LM1, contribui perceptivelmente na sonoridade de pianos, principalmente nas notas mais graves. Para uma corda sem enrolamento, a freqüência de LM1 é aproximadamente igual a $2500/l$ Hz [17]. Esta freqüência normalmente não coincide com nenhuma das freqüências da série de modos transversais da corda, o que provoca um efeito dissonante no som. Para contornar este problema, muitos fabricantes controlam as freqüências dos modos LM1 alterando as características dos enrolamentos utilizados sobre cada corda. Conseguem assim obter freqüências consonantes com o

pitch de cada nota.

1.9 Histórico

Um sistema desenvolvido para a identificação de notas de sinais polifônicos de violão foi apresentado por BONNET e LEFEBVRE [18]. Neste, a análise é realizada sobre trechos segmentados de sinais, correpondentes à sustentação das notas. Os autores desenvolveram uma heurística para identificar no espectro freqüencial, normalizado e suavizado, os picos correspondentes às freqüências fundamentais de notas musicais.

GAGNON *et al.* [19], propuseram um método de auxílio ao reconhecimento de acordes em sinais de instrumentos solo, em especial de violão. O método, com base em redes neurais e representação da distribuição energética dos sinais sobre a escala Bark [20], tem como objetivo indicar o número de cordas usadas na geração do acorde e a posição sobre o braço do instrumento em que as notas foram tocadas.

A aplicação de redes neurais *feed-forward* na identificação de notas em sinais polifônicos foi estudada por MAROLT [21], que desenvolveu um sistema para transcrição de gravações de piano que envolve, além de redes neurais, um modelo auditivo e redes adaptativas de osciladores usadas no rastreamento de parciais.

SZCZUPAK *et al.* [22] apresentaram um estudo sobre a identificação notas musicais em registros polifônicos de violão através de redes neurais. Neste estudo foram desenvolvidas seis redes neurais projetadas para a análise de espectros obtidos através de uma transformada de Q constante. Cada rede foi desenvolvida para a identificação de notas em registros com graus diferentes de polifonia.

KLAPURI [23] desenvolveu um método para estimar freqüências fundamentais de sinais polifônicos sem restrições em relação aos instrumentos presentes na gravações. O método, com base em modelos computacionais de percepção de *pitch*, é utilizado para estimar, uma a uma, as freqüências fundamentais das notas presentes no sinal. Para cada freqüência fundamental estimada, se busca subtrair do espectro do sinal a contribução de parte das parciais relacionadas a esta freqüência. O processo é então repetido iterativamente no sinal residual.

RYYNÄNEN e KLAPURI [24] associaram o método desenvovido por KLAPURI [23] a um sistema completo de transcrição musical que utiliza um modelo

probabilístico, descrito por *hidden Markov models* [25], para a análise das notas ao longo da duração dos sinais. Este sistema foi projetado para transcrever gravações de instrumentos com *pitch* definido, incluindo misturas de instrumentos diferentes, porém com extensões restritas à região que compreende as notas F 1 e B♭6.

POLINER e ELLIS [26] desenvolveram um sistema para transcrição musical de gravações de piano que tem como base classificadores do tipo máquina de vetor de suporte [27], treinados com representações espectrais. O sistema trata o problema de identificação das notas como um grupo de classificações binárias. São utilizados 87 classificadores OVA (*one-versus-all*), cada um para detecção de uma nota diferente.

1.10 Proposta da Dissertação

Comumente, sinais discretos são representados no domínio da freqüência através da DFT, com as componentes resultantes distribuídas ao longo de uma escala linear de freqüências. Assim, as oitavas mais altas do espectro freqüencial são representadas com mais componentes que as oitavas mais baixas.

Já nas escalas musicais de temperamento igual, as freqüências de cada nota são dispostas em uma progressão geométrica com razão $2^{1/12}$. Conseqüentemente, quando a gravação de um instrumento afinado em escala de temperamento igual é analisada usando-se a DFT, a quantidade de linhas espectrais em torno do conjunto de parciais de notas mais graves é menor do que em torno do conjunto de parciais de notas mais agudas.

Como alternativa para equalizar a análise de diferentes notas, pode-se utilizar a CQT (*constant-Q transform*) [28], uma transformada espectral discreta com seletividade constante e freqüências espaçadas em progressão geométrica, assim como as das notas de escalas de temperamento igual.

Para identificar as notas musicais presentes em gravações de violão, são propostos métodos de classificação com base em redes neurais *feed-forward* de múltiplas camadas, treinadas com representações freqüenciais obtidas pela CQT. Este tipo de rede é apropriado para tarefas de classificação que envolvem padrões não linearmente separáveis [29].

O projeto das redes foi realizado explorando propriedades da CQT e carac-

terísticas do violão. Busca-se abordar aspectos de execução musical que podem ser observados em registros do instrumento, como variações na acentuação das notas (dinâmica) e a análise de segmentos de sinal compostos por notas soando durante diferentes períodos do modelo ADSR.

Os métodos propostos foram desenvolvidos e testados utilizando sinais gerados computacionalmente pela combinação de trechos de registros reais de notas musicais de violão. Foram utilizados registros com três níveis diferentes de dinâmica, *piano*, *mezzo* e *forte*. A escolha dos segmentos utilizados de cada registro foi realizada de acordo com uma seqüência de quatro objetivos:

1. Identificar notas em combinações de registros com dinâmica *mezzo* a partir de segmentos extraídos aproximadamente do período de sustentação de cada nota.

2. Identificar notas em combinações de registros com dinâmica *mezzo* (exceto por um, com dinâmica *forte*) a partir de segmentos extraídos aproximadamente do período de sustentação de cada nota.

3. Identificar notas em combinações de registros com dinâmica *mezzo* a partir de três possibilidades de segmentação: todos os segmentos extraídos aproximadamente do período que compreende o ataque e decaimento, todos os segmentos extraídos aproximadamente do período de sustentação e todos os segmentos extraídos aproximadamente do período de liberação.

4. Identificar notas em combinações de registros que têm, independentemente, um entre três níveis de dinâmica (*piano*, *mezzo* ou *forte*) a partir de segmentos extraídos, independentemente, de um entre três períodos (aproximadamente do período que compreende o ataque e o decaimento, aproximadamente do período de sustentação ou aproximadamente do período de liberação).

Também são apresentadas adaptações dos métodos voltadas para a identificação de notas musicais em sinais de piano. Estas adaptações foram aplicados para identificar notas em combinações de registros que têm, independentemente, um entre três níveis de dinâmica (*piano*, *mezzo* ou *forte*) a partir de segmentos extraídos, independentemente, de um entre três períodos (aproximadamente do período que

compreende o ataque e o decaimento, aproximadamente do período de sustentação ou aproximadamente do período de liberação). O objetivo é obter material para comparação de resultados com outros métodos de identificação de notas em registros polifônicos, comumente desenvolvidos para piano solo porém escassos para violão.

Capítulo 2

A Transformada de Q Constante

2.1 Descrição

A CQT é uma transformada espectral com seletividade constante e componentes definidas sobre uma escala de freqüências em progressão geométrica:

$$f[k_{\mathrm{cq}}] = q^{k_{\mathrm{cq}}} f_{\mathrm{min}}, \qquad k_{\mathrm{cq}} = 0, 1, ..., k_{\mathrm{max}} \ , \tag{2.1}$$

onde:

f_{min} = freqüência mínima escolhida para a análise,

f_{s} = freqüência de amostragem do sinal e

$2f[k_{\mathrm{max}}] < f_{\mathrm{s}}$.

Para facilitar a análise de sinais musicais, esta escala pode ser gerada com:

$$q = 2^{\frac{1}{12\beta}}, \qquad \beta \in \{1, 2, 3, ...\} \, . \tag{2.2}$$

O fator β define a resolução espectral em frações de semitom. Quanto maior o valor de β, maior a resolução e a seletividade

$$Q = \frac{f[k_{\mathrm{cq}}]}{qf[k_{\mathrm{cq}}] - f[k_{\mathrm{cq}}]} = \frac{1}{q-1} \tag{2.3}$$

da transformada. Por exemplo, com $\beta = 1$ e f_{min} coincidente com o *pitch* de uma nota musical, os valores de $f[k_{\mathrm{cq}}]$ coincidem com os *pitches* de uma seqüência de notas espaçadas por um intervalo de semitom[1].

[1] A coincidência dos valores de $f[k_{\mathrm{cq}}]$ com os *pitches* de uma seqüência de notas não significa

O espectro X_{cq} da CQT de $x[n]$ é dado por:

$$X_{cq}[k_{\mathrm{cq}}] = \frac{1}{N[k_{\mathrm{cq}}]} \sum_{n=0}^{N[k_{\mathrm{cq}}]-1} w[n, k_{\mathrm{cq}}] x[n] e^{-j2\pi \frac{Q}{N[k_{\mathrm{cq}}]} n}, \tag{2.4}$$

$$N[k_{\mathrm{cq}}] = \frac{f_s Q}{f[k_{\mathrm{cq}}]}, \tag{2.5}$$

sendo $w[n, k_{\mathrm{cq}}]$ uma função-janela de comprimento $N[k_{\mathrm{cq}}]$.

A Figura 2.1 contém um esquema que relaciona *pitches* em uma escala de temperamento igual (representados por teclas de piano) às freqüências das componentes de uma CQT e de uma DFT. Neste exemplo, ambas as transformadas têm o mesmo número de componentes sobre a faixa de freqüências representada. Na CQT, neste caso com $\beta = 1$, a densidade de componentes por elementos da escala permanece constante. Na DFT a densidade de componentes por elementos cresce com o incremento da freqüência.

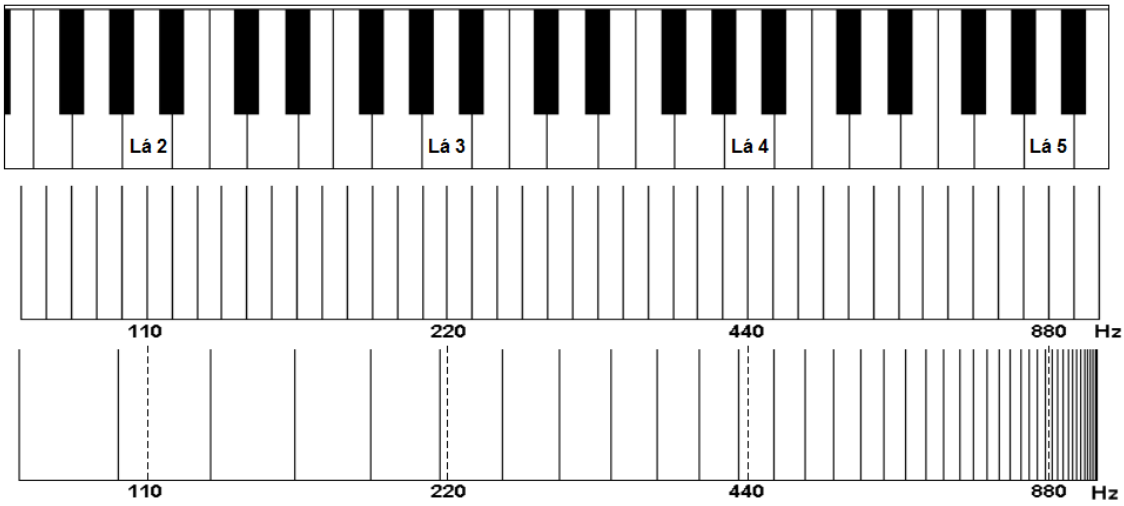

Figura 2.1: Esquema que relaciona *pitches* em uma escala de temperamento igual, representados por teclas de piano na parte superior da figura, às freqüências das componentes de uma CQT e de uma DFT, representadas no meio e na parte inferior da figura, respectivamente.

---

que exista coincidência dos valores de $f[k_{\mathrm{cq}}]$ com as freqüências fundamentais das notas. Devido à inarmonicidade presente em instrumentos reais, discutida no capítulo introdutório, as freqüências fundamentais das notas de um instrumento afinado em temperamento igual não seguem uma progressão geométrica exata.

2.2 Algoritmo Rápido

Nesta tese a CQT foi implementada através de um algoritmo rápido [30], com base no algoritmo FFT.

Definindo um *kernel* temporal κ para cada k_{cq}, na forma:

$$\kappa[n, k_{\mathrm{cq}}] = w[n, k_{\mathrm{cq}}] e^{j2\pi \frac{f[k_{\mathrm{cq}}]}{f_{\mathrm{s}}} n}, \tag{2.6}$$

segue

$$X_{cq}[k_{\mathrm{cq}}] = \sum_{n=0}^{N-1} x[n] \kappa^*[n, k_{\mathrm{cq}}] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] K^*[k, k_{\mathrm{cq}}], \tag{2.7}$$

onde

$$N = \frac{f_{\mathrm{s}} Q}{f[0]} \tag{2.8}$$

e

$$K[k, k_{\mathrm{cq}}] = \sum_{n=0}^{N-1} w \left[n - \left(\frac{N}{2} - \frac{N(k_{\mathrm{cq}})}{2} \right), k_{\mathrm{cq}} \right] e^{j2\pi \frac{f[k_{\mathrm{cq}}]}{f_{\mathrm{s}}} \left(n - \frac{N}{2}\right)} e^{-j2\pi \frac{kn}{N}} \tag{2.9}$$

é o *kernel* freqüencial, dado pela DFT de $\kappa \left[n - \frac{N}{2}, k_{\mathrm{cq}} \right]$. Na Equação (2.9), a função w tem o mesmo número de amostras N para todo k_{cq}, porém seus valores fora de cada intervalo $\left(\frac{N}{2} - \frac{N[k_{\mathrm{cq}}]}{2}, \frac{N}{2} + \frac{N[k_{\mathrm{cq}}]}{2} \right)$ são iguais a zero. Dentro deste intervalo, w é uma janela de ponderação. Nesta tese foram utilizadas janelas de Hamming. Assim, $w[n, k_{\mathrm{cq}}] =$

$$\begin{cases} 0{,}54 - 0{,}46 \cos \frac{2\pi}{N[k_{\mathrm{cq}}]} \left(n - \left(\frac{N}{2} - \frac{N[k_{\mathrm{cq}}]}{2} \right) \right), & n \in \left\{ \frac{N}{2} - \frac{N[k_{\mathrm{cq}}]}{2}, \frac{N}{2} + \frac{N[k_{\mathrm{cq}}]}{2} \right\} \\ w[n, k_{\mathrm{cq}}] = 0, & n \notin \left\{ \frac{N}{2} - \frac{N[k_{\mathrm{cq}}]}{2}, \frac{N}{2} + \frac{N[k_{\mathrm{cq}}]}{2} \right\} \end{cases}$$

Como os *kernels* temporais são seqüências simétricas conjugadas ($\kappa[n, k_{\mathrm{cq}}] = \kappa^*[-n, k_{\mathrm{cq}}]$), os *kernels* freqüenciais $K[k_{\mathrm{cq}}]$ são reais [31]. Cada *kernel* freqüencial apresenta valores significativos apenas para uma faixa concentrada de valores de k. Considerando nulos os valores muito pequenos, pode-se reduzir drasticamente o número de multiplicações realizadas, obtendo assim o algoritmo rápido.

2.3 Transposição

Transpor um acorde significa mudar suas notas sem alterar seus intervalos. Para um vetor contendo os valores absolutos dos elementos de uma CQT, a transpo-

sição corresponde a um deslocamento igual de cada um destes elementos pela escala de freqüências $f[k_{\mathrm{cq}}]$.

Sendo $X_{\mathrm{cq}}[k_{\mathrm{cq}}]$ a CQT de $x[n]$, $T_{\mathrm{cq}}[k_{\mathrm{cq}}]$ a CQT de $x[n]$ transposta por q^{k_d} e abs$(\cdot)$ uma função que, aplicada a um vetor, retorna os valores absolutos de cada um de seus elementos, então:

$$\mathrm{abs}(T_{\mathrm{cq}}[k_{\mathrm{cq}}]) = \mathrm{abs}(X_{\mathrm{cq}}[k_{\mathrm{cq}} - k_d]). \tag{2.10}$$

Para ilustrar esta propriedade, nas Figuras 2.2 e 2.3 são mostrados gráficos dos valores absolutos dos elementos de duas transformadas realizadas com $\beta = 1$.

A Figura 2.2 contém o gráfico da CQT de uma composição de senóides com freqüências da série harmônica de Lá 2 (freqüência fundamental=110,00 Hz). A Figura 2.3 contém o gráfico da CQT de uma composição de senóides com freqüências da série harmônica de Dó♯4 (freqüência fundamental≈277,18 Hz). Ambas as composições foram geradas artificialmente. Cada uma é formada pela soma de cinco senóides com amplitudes ponderadas, da mais baixa à mais alta, por 1; 1/2; 1/4; 1/8 e 1/16, contaminadas por ruído aditivo Gaussiano branco (SNR=10 dB).

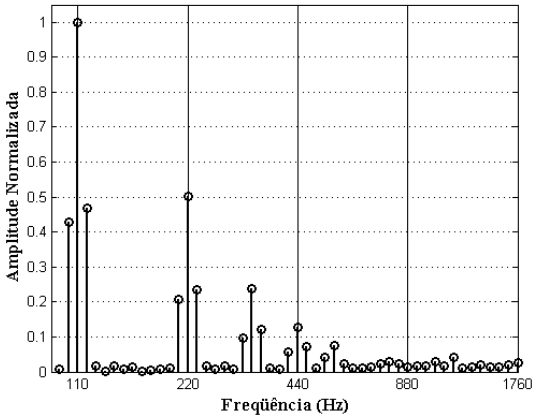

Figura 2.2: CQT de harmônicos de Lá 2.

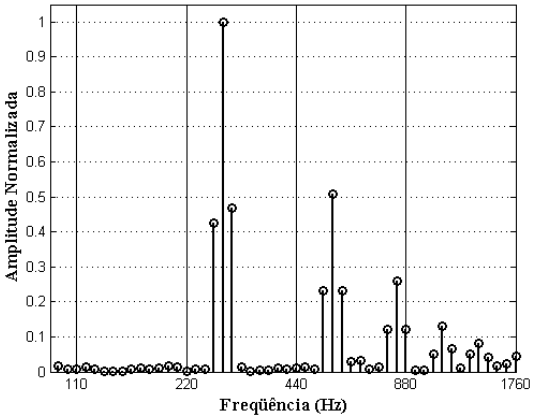

Figura 2.3: CQT de harmônicos de Dó♯4.

Da transposição 16 semitons acima de um acorde que contém a nota Lá 2, obtém-se um acorde que contém a nota Dó♯4. Do mesmo modo, do deslocamento 16β componentes acima do espectro CQT de uma nota Lá 2, obtém-se o espectro de uma nota Dó♯4.

2.4 Análise Tempo-Freqüência

O centro da análise de todos os intervalos de $N[k_{\mathrm{cq}}]$ amostras utilizadas para calcular a CQT é o mesmo para todas as k_{cq} componentes; porém, o número de amostras utilizadas depende do valor de cada componente analisada na Equação (2.5). Quanto menor é a componente k_{cq}, maior é o valor de $N[k_{\mathrm{cq}}]$. Assim, em uma análise tempo-freqüência composta por uma seqüência de CQTs tomadas ao longo da duração de um sinal, o número de amostras sobrepostas depende tanto da quantidade de amostras h (hop) entre centros consecutivos de análise, quanto da componente k_{cq} analisada.

Para que todas as amostras do sinal sejam analisadas para cada componente k_{cq}, é necessário estabelecer um passo h com comprimento máximo igual ao comprimento do menor $N[k_{\mathrm{cq}}]$. Assim,

$$h \leq \frac{f_{\mathrm{s}}\, Q}{f[k_{\mathrm{max}}]}. \tag{2.11}$$

Com esta escolha, para grandes extensões de freqüências ($f[k_{\mathrm{max}}] >> f[0]$), o passo h será bem menor que o comprimento do maior intervalo analisado ($h << N[0]$). Além do elevado custo computacional decorrente do h reduzido, haverá grande sobreposição entre intervalos $N[k_{\mathrm{cq}}]$ consecutivos para as componentes de freqüências mais baixas, resultando em uma análise redundante desta faixa do espectro. Como alternativa, pode-se optar por um passo de comprimento intermediário:

$$\frac{f_{\mathrm{s}}\, Q}{f[k_{\mathrm{max}}]} < h < \frac{f_{\mathrm{s}}\, Q}{f[0]}. \tag{2.12}$$

Neste caso, haverá amostras nunca analisadas durante os cálculos das componentes mais elevadas do espectro. Como conseqüência, eventos transitórios no sinal analisado com energia significativa nesta faixa podem não ser satisfatoriamente descritos.

2.5 Estacionariedade

A dependência entre o número de amostras $N[k_{\text{cq}}]$ e a freqüência de cada componente $f[k_{\text{cq}}]$ (Equação (2.5)) não é condição suficiente para garantir a seletividade constante da CQT. Também é necessário que as componentes freqüenciais do sinal permaneçam estacionárias ao longo de cada janela $w[n, k_{\text{cq}}]$. Isto pode não se verificar em sinais de música reais, principalmente se a análise for realizada sobre componentes de baixa freqüência.

Por exemplo: a duração do intervalo de análise da CQT, com $\beta = 1$, para uma componente centrada em $f = 27,5$ Hz (*pitch* da nota Lá 0) é aproximadamente 612 ms. Tipicamente, um sinal de áudio real pode ser considerado aproximadamente estacionário por cerca de 20 ms. Assim, as análises sobre componentes de baixa freqüência podem acabar sendo realizadas sobre períodos não-estacionários de sinal.

Capítulo 3

Redes Neurais *Feed-Forward* de Múltiplas Camadas

3.1 Redes Neurais

Redes neurais artificiais são estruturas computacionais para processamento de informação inspiradas no funcionamento cerebral. São compostas por combinações de estruturas computacionais básicas, os neurônios artificiais, por sua vez inspiradas no funcionamento dos neurônios biológicos. Uma rede neural artificial não emula todo o funcionamento cerebral; em vez disto, de acordo com seu projeto, modela apenas alguns de seus mecanismos.

Entre os mecanismos comumente modelados estão: o processamento em paralelo de informação, a capacidade de aprender (para as redes neurais artificiais, 'capacidade de aprender' pode ser entendida como a capacidade de adaptação de sua arquitetura e de seus parâmetros livres para melhor desempenhar uma determinada tarefa [32]) e a capacidade de generalização (uma rede neural projetada para reconhecimento de padrões pode, após um processo de treinamento, tornar-se imune a pequenas variações dos sinais de entrada, sendo assim apropriada para processamento de sinais com ruído ou distorção [33]).

O projeto de uma rede neural artificial (bem como dos neurônios artificiais) depende do objetivo do processamento. Entre os objetivos típicos estão: reconhecimento de padrões, aproximação de funções e clusterização. As redes do tipo *feed-forward* de múltiplas camadas podem ser treinadas, através do algoritmo *backpropa-*

gation (descrito na Seção 3.4), para tarefas de reconhecimento de padrões, aplicação de interesse nesta tese. O algoritmo *backpropagation* descreve como modificar os pesos sinápticos da rede utilizando sinais de entrada para os quais as saídas desejadas são conhecidas. A modificação dos pesos sinápticos visa a minimizar uma medida de erro entre as saídas da rede e as saídas desejadas (que, no caso das aplicações de reconhecimento, identificam os padrões dos sinais de entrada). Uma rede treinada para reconhecimento de padrões deve, na presença de um sinal de entrada pertencente a algum dos padrões treinados, gerar uma saída coerente com este padrão, mesmo que o sinal seja inédito.

3.2 Os Neurônios Artificiais

O modelo de neurônio[1] de uma rede *feed-forward* de múltiplas camadas é formado por [27]: um conjunto de sinapses que conectam o vetor de entrada ao neurônio, associando cada elemento do vetor a um fator multiplicador (peso sináptico); um somador que opera sobre os elementos do vetor de entrada (ponderados pelos respectivos pesos sinápticos) e sobre um elemento de polarização; e uma função de ativação, que recebe como argumento o campo do neurônio (a saída do somador).

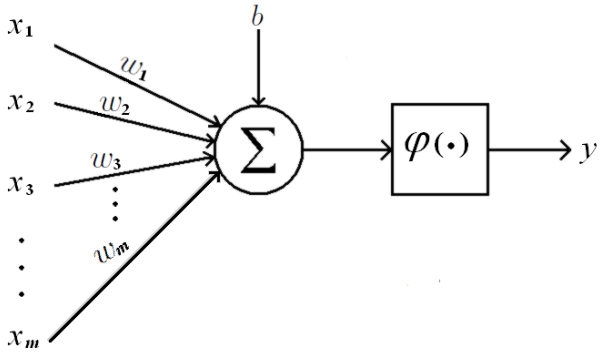

Figura 3.1: Diagrama de um modelo de neurônio artificial.

A Figura 3.1 contém o diagrama de um neurônio de uma rede *feed-forward* de múltiplas camadas. O neurônio, com vetor de entrada $(x_1\ x_2\ ...\ x_m)^{\mathrm{T}}$, tem pesos

[1] A partir deste ponto, por brevidade, a palavra "neurônio" substitui "neurônio artificial". Do mesmo modo, "rede neural" substituirá adiante "rede neural artificial".

sinápticos $\{w_1 \ ... \ w_m\}$. Assim, seu campo u é dado por

$$u = \sum_{p=1}^{m} (x_p w_p) + b, \tag{3.1}$$

onde w_p é o peso sináptico associado ao elemento x_p e b é o elemento de polarização (ou *bias*) do campo.

O termo b pode ser modelado como o produto entre uma entrada $x_0 = 1$ e um peso sináptico $w_0 = b$. Assim u pode ser reescrito como

$$u = \sum_{p=0}^{m} x_p w_p, \tag{3.2}$$

e a saída y do neurônio pode ser escrita como

$$y = \varphi \left(\sum_{p=0}^{m} x_p w_p \right) = \varphi \left(\mathbf{w}^{\mathrm{T}} \mathbf{x} \right), \tag{3.3}$$

onde $\mathbf{w} = (w_0 \ w_1 \ ... \ w_m)^{\mathrm{T}}$ e $\mathbf{x} = (x_0 \ x_1 \ ... \ x_m)^{\mathrm{T}}$.

O neurônio normalmente é representado de uma forma simplificada, mostrada na Figura 3.2.

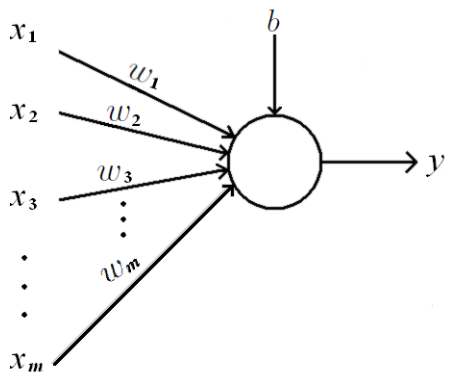

Figura 3.2: Diagrama simplificado de neurônio de uma rede *feed-forward*.

A função de ativação φ é usualmente uma função não-linear suave. Isto permite que, em aplicações de reconhecimento de padrões, se obtenham fronteiras de decisão suaves entre classes. Para realizar o treinamento da rede utilizando o algoritmo *back-propagation* é necessário que a função de ativação seja diferenciável. Funções sigmoidais como a logística (Equação 3.4) e a tangente hiperbólica são opções comuns que atendem estas condições. Elas são funções monotônicas que apresentam comportamento aproximadamente linear para argumentos com pequeno

valor absoluto, porém suas saídas se aproximam assintoticamente de um limite superior ou inferior conforme o argumento cresce ou decresce. Os limites da função logística e da função tangente hiperbólica são, respectivamente, {0,1} e {-1,1}. Em geral, quando se utilizam funções sigmoidais, as saídas desejadas são compostas com valores pertencentes aos limites das funções de ativação utilizadas nos neurônios de saída da rede. A função logística é dada por

$$\varphi(u) = \frac{1}{1 + \exp^{-\sigma u}}, \qquad \sigma > 0, \tag{3.4}$$

onde o parâmetro σ controla a inclinação da função ao longo de sua imagem.

As Figuras 3.3 e 3.4 contêm as respostas das funções logística e tangente hiperbólica para valores de entrada dentro da faixa $[-6, 6]$.

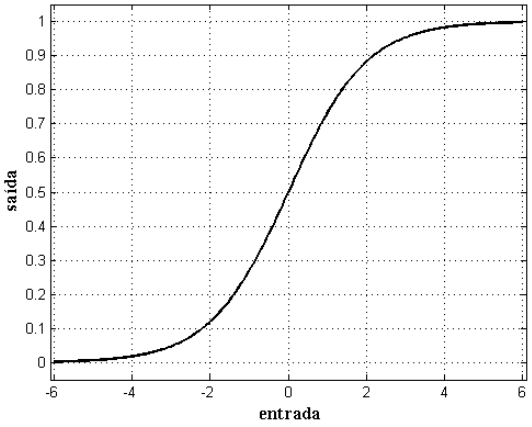

Figura 3.3: Curva de respostas da função logística com $\sigma = 1$.

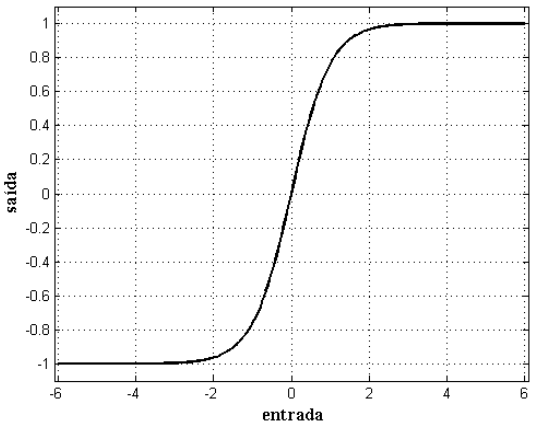

Figura 3.4: Curva de respostas da função tangente hiperbólica.

3.3 A Organização em Camadas

As redes *feed-forward* podem ter uma ou mais camadas de neurônios. Nelas não existem ligações de realimentação entre neurônios de diferentes camadas, nem ligações entre neurônios de uma mesma camada. Cada neurônio recebe como entrada apenas saídas de neurônios de camadas precedentes ou, no caso da primeira camada, do vetor de entrada da rede. Na configuração mais comum, a rede é organizada em camadas ligadas em cascata. As saídas dos neurônios de uma camada servem como entradas para os neurônios da camada seguinte. A 1^a camada de neurônios recebe

o vetor de entrada da rede, a 2^a camada recebe o vetor composto pelas saídas da 1^a camada, a 3^a camada — se existir — recebe o vetor composto pelas saídas da 2^a camada, e assim por diante.

As redes são chamadas totalmente conectadas se todos os neurônios de uma camada qualquer tiverem ligações sinápticas com as saídas de todos os neurônios da camada anterior (ou, no caso da 1^a camada, se todos os neurônios tiverem conexões sinápticas com todos os elementos do vetor de entrada da rede). Todas as camadas de neurônios, exceto a camada de saída, são chamadas de camadas ocultas.

A Figura 3.5 contém a representação de uma rede *feed-forward* de duas camadas com vetor de entrada $(x_1\ x_2\ ...\ x_m)^{\mathrm{T}}$. A rede é totalmente conectada, com três neurônios na primeira camada (a camada oculta) e dois neurônios na segunda (a camada de saída)[2].

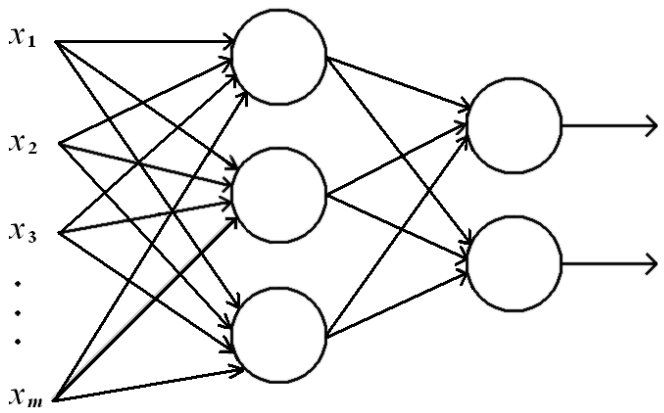

Figura 3.5: Rede neural *feed-forward* de duas camadas. Para simplificar o desenho, os pesos sinápticos e elementos de polarização dos neurônios não estão representados.

Durante o treinamento supervisionado, através de uma transformação não-linear dos dados de entrada para um novo espaço, os neurônios das camadas ocultas extraem progressivamente dos vetores de entrada as características mais significativas para a separação das classes [27]. Neste novo espaço, as classes de interesse podem ser mais facilmente separadas entre si do que no espaço original de entrada.

[2]Alguns autores se referem ao vetor de entrada como a 1^a camada da rede. Neste caso, uma rede como a da Figura (3.5) seria classificada como *feed-forward* de três camadas. Nesta tese adota-se a convenção em que o vetor de entrada é nomeado 'camada 0' e as redes são nomeadas de acordo com o número de camadas de neurônios, apenas.

3.4 Algoritmo *Backpropagation*

Para realizar o treinamento por *backpropagation* é necessário ter um conjunto de vetores de entrada para os quais se conhecem os vetores desejados como saídas da rede. Em uma tarefa de reconhecimento de padrões, todos os vetores de entrada pertencentes a uma mesma classe são usualmente associados ao mesmo vetor-objetivo.

O algoritmo é realizado em iterações sucessivas, cada uma composta por duas etapas: a propagação adiante do sinal de entrada, quando, de acordo com o vetor de entrada apresentado, são calculadas as saídas de cada camada da rede para a configuração corrente de pesos sinápticos; e a retropropagação do erro, quando são calculados os ajustes dos pesos sinápticos em função de uma medida de erro entre o vetor de saída da rede e o vetor-objetivo associado ao vetor de entrada apresentado. Considerando uma rede com camadas de neurônios numeradas $c = 1, ..., s$ (sendo s a camada de saída), nesta etapa inicialmente são calculados os ajustes dos pesos sinápticos da camada s, em seguida são calculados os ajustes dos pesos da camada $s-1$, e assim por diante até serem calculados os ajustes dos pesos da camada 1.

Antes de realizar o treinamento é aconselhável que todos os pesos sinápticos sejam inicializados com valores escolhidos randomicamente e pequenos o suficiente para que não ocorram saturações em neurônios na iteração inicial [33]. A saturação de um neurônio com função de ativação sigmoidal ocorre quando sua saída se aproxima de um de seus limites e a derivada de sua função de ativação, em relação ao campo do neurônio, se aproxima de zero.

Na etapa de propagação adiante do sinal de entrada, para cada vetor de entrada $(x_1 \ ... \ x_m)^{\mathrm{T}}$ são calculadas as saídas de cada neurônio de acordo com a configuração corrente de pesos sinápticos.

Sendo $w_{c,pq}$ o peso sináptico do q-ésimo neurônio da camada c que pondera a p-ésima saída da camada anterior $y_{(c-1),p}$, a saída de um neurônio r de uma camada $(c{=}v)$ é dada por:

$$y_{v,r} = \varphi\left(\mathbf{w}_v^{\mathrm{T}}\mathbf{y}_{(v-1)}\right), \tag{3.5}$$

onde

$\mathbf{w}_v = (w_{v,0r} \ w_{v,1r} \ ... \ w_{v,mr})^{\mathrm{T}}$,

$w_{v,0r} = b_{v,r}$ (elemento de polarização do r-ésimo neurônio da v-ésima camada),

m = número de elementos de saída na camada $v-1$,

$\mathbf{y}_{(v-1)} = (y_{(v-1),0}\ \ y_{(v-1),1}\ \ \cdots\ \ y_{(v-1),m})^{\mathrm{T}}$ e

$y_{(v-1),0} = 1$.

A camada anterior à primeira camada de neurônios ($c=1$) é o vetor de entrada (a camada 0), assim a saída de um neurônio p da 1^a camada é dada por:

$$y_{1,p} = \varphi\left(\mathbf{w}_1^{\mathrm{T}}\mathbf{x}\right), \tag{3.6}$$

onde

$\mathbf{w}_1 = (w_{1,0p}\ w_{1,1p}\ ...\ w_{1,mp})^{\mathrm{T}}$,

$w_{1,0p} = b_{1,p}$ (elemento de polarização do p-ésimo neurônio da 1^a camada),

$\mathbf{x} = (x_0\ x_1\ ...\ x_m)^{\mathrm{T}}$ e

$x_0 = 1$.

A saída de um neurônio q da segunda camada ($c=2$) é dada por:

$$y_{2,q} = \varphi\left(\mathbf{w}_2^{\mathrm{T}}\mathbf{y}_1\right), \tag{3.7}$$

onde

$\mathbf{w}_2 = (w_{2,0q}\ w_{2,1q}\ ...\ w_{2,mq})^{\mathrm{T}}$,

$w_{2,0q} = b_{2,q}$ (elemento de polarização do q-ésimo neurônio da 2^a camada),

$\mathbf{y}_1 = (y_{1,0}\ y_{1,1}\ ...\ y_{1,m})^{\mathrm{T}}$ e

$y_{1,0} = 1$.

Na retropropagação do erro, sendo $\mathbf{y}_s = (y_{s,1}\ \ y_{s,2}\ \ ...\ \ y_{s,j})^{\mathrm{T}}$ o vetor de saída da rede, $\mathbf{d} = (d_1\ d_2\ ...\ d_j)^{\mathrm{T}}$ o vetor-objetivo, $e_q = d_q - y_{s,q}$ (diferença entre o q-ésimo elemento do vetor-objetivo e o q-ésimo elemento do vetor de saída) e a função de custo E, a minimizar, igual à soma dos quadrados das diferenças e_q, isto é,

$$E = \sum_{q=1}^{j} e_q^2 = \sum_{q=1}^{j}(d_q - y_{s,q})^2, \tag{3.8}$$

para ajustar o peso $w_{c,pq}$ é necessário produzir um ajuste $\Delta w_{c,pq}$ no sentido de descida do gradiente da superfície de custo em relação ao espaço de pesos sinápticos,

$$\Delta w_{c,pq} = -\alpha\frac{\partial E}{\partial w_{c,pq}}. \tag{3.9}$$

A taxa α é utilizada para controlar a evolução do processo de treinamento. Uma taxa muito elevada pode tornar o processo de treinamento instável. Uma taxa muito

pequena pode tornar o processo muito lento. A escolha de um valor ótimo para esta constante depende do problema tratado.

Para ajustar um neurônio da camada de saída ($w_{c,pq} = w_{s,pq}$), é necessário exprimir $\dfrac{\partial E}{\partial w_{s,pq}}$ em função de valores conhecidos (calculados na etapa de propagação adiante do sinal de entrada). Utilizando a regra da cadeia pode-se escrever:

$$\frac{\partial E}{\partial w_{s,pq}} = \frac{\partial E}{\partial e_q}\,\frac{\partial e_q}{\partial y_{s,q}}\,\frac{\partial y_{s,q}}{\partial u_{s,q}}\,\frac{\partial u_{s,q}}{\partial w_{s,pq}}. \tag{3.10}$$

A seguir são descritos, em função dos valores calculados na etapa de propagação adiante do sinal de entrada, os fatores à direita da Equação (3.10).

$$\frac{\partial E}{\partial e_q} = \frac{\partial \left(\sum_{q=1}^{j} e_q^2 \right)}{\partial e_q} = 2e_q, \tag{3.11}$$

$$\frac{\partial e_q}{\partial y_{s,q}} = \frac{\partial (d_q - y_{s,q})}{\partial y_{s,q}} = -1, \tag{3.12}$$

$$\frac{\partial y_{s,q}}{\partial u_{s,q}} = \frac{\partial \varphi(u_{s,q})}{\partial u_{s,q}}, \tag{3.13}$$

$$\frac{\partial u_{s,q}}{\partial w_{s,pq}} = \frac{\partial \left(\sum_{p} y_{(s-1),p}\ w_{s,pq} \right)}{\partial w_{s,pq}} = y_{(s-1),p}\,. \tag{3.14}$$

Substituindo as soluções das Equações (3.11), (3.12), (3.13) e (3.14) na Equação (3.10), segue:

$$\frac{\partial E}{\partial w_{s,pq}} = -2e_q \frac{\partial \varphi(u_{s,q})}{\partial u_{s,q}} y_{(s-1),p}\,. \tag{3.15}$$

O produto $e_q \dfrac{\partial \varphi(u_{s,q})}{\partial u_{s,q}}$, parte da Equação (3.15), pode ser reescrito em função de $\dfrac{\partial E}{\partial u_{s,q}}$. Utilizando-se o resultado da Equação (3.11) pode-se escrever:

$$\delta_{s,q} = e_q \frac{\partial \varphi(u_{s,q})}{\partial u_{s,q}} = \frac{1}{2}\,\frac{\partial E}{\partial e_q}\,\frac{\partial \varphi(u_{s,q})}{\partial u_{s,q}}. \tag{3.16}$$

Pela regra da cadeia,

$$\delta_{s,q} = \frac{1}{2}\,\frac{\partial E}{\partial y_{s,q}}\,\frac{\partial y_{s,q}}{\partial e_q}\,\frac{\partial \varphi(u_{s,q})}{\partial u_{s,q}}. \tag{3.17}$$

Substituindo o resultado da Equação (3.12) na Equação (3.17),

$$\delta_{s,q} = -\frac{1}{2}\,\frac{\partial E}{\partial y_{s,q}}\,\frac{\partial \varphi(u_{s,q})}{\partial u_{s,q}}. \tag{3.18}$$

Como $\varphi(u_{s,q}) = y_{s,q}$,

$$\delta_{s,q} = -\frac{1}{2}\,\frac{\partial E}{\partial y_{s,q}}\,\frac{\partial y_{s,q}}{\partial u_{s,q}}, \tag{3.19}$$

$$\delta_{s,q} = -\frac{1}{2}\,\frac{\partial E}{\partial u_{s,q}}. \tag{3.20}$$

Substituindo o resultado da Equação (3.20) na Equação (3.15):

$$\frac{\partial E}{\partial w_{s,pq}} = -2\,\delta_{s,q}\,y_{(s-1),p}\,. \tag{3.21}$$

A regra de atualização para pesos sinápticos da camada de saída ($c{=}s$) pode ser obtida através da substituição do resultado da Equação (3.21) na Equação (3.9),

$$\Delta w_{s,pq} = 2\,\alpha\,\delta_{s,q}\,y_{(s-1),p}\,. \tag{3.22}$$

Os pesos sinápticos da camada $c{=}s-1$ podem ser corrigidos após os ajustes da camada de saída ($c{=}s$). Para ajustar o peso sináptico $w_{(s-1),rp}$ (o peso sináptico do p-ésimo neurônio da camada $s-1$ e que pondera $y_{(s-2),r}$, a r-ésima saída da camada $s-2$)[3] é necessário, analogamente ao procedimento utilizado para ajustar o peso $w_{s,pq}$, produzir um ajuste $\Delta w_{(s-1),rp}$ no sentido de descida do gradiente da superfície de custo em relação ao espaço de pesos sinápticos:

$$\Delta w_{(s-1),rp} = -\alpha\frac{\partial E}{\partial w_{(s-1),rp}}. \tag{3.23}$$

Assim como para ajustar um neurônio q da camada de saída é necessário utilizar a derivada parcial da função de custo em relação ao campo deste neurônio $\left(\dfrac{\partial E}{\partial u_{s,q}}\right)$ — Equações (3.19) e (3.20) —, para calcular o ajuste de um neurônio p da camada $s-1$, é necessário utilizar a derivada parcial da função de custo em relação ao campo $u_{(s-1),p}$, $\left(\dfrac{\partial E}{\partial u_{(s-1),p}}\right)$. Pela regra da cadeia,

$$\Delta w_{(s-1),rp} = -\alpha\frac{\partial E}{\partial u_{(s-1),p}}\,\frac{\partial u_{(s-1),p}}{\partial w_{(s-1),rp}}. \tag{3.24}$$

[3]Se a rede tiver duas camadas de neurônios, então $y_{(s-2),r} = x_r$ (o r-ésimo elemento do sinal de entrada da rede).

Como

$$\frac{\partial u_{(s-1),p}}{\partial w_{(s-1),rp}} = \frac{\partial \left(\sum_r y_{(s-2),r} \; w_{(s-1),rp} \right)}{\partial w_{(s-1),rp}} = y_{(s-2),r} \; , \tag{3.25}$$

então:

$$\Delta w_{(s-1),rp} = -\alpha \frac{\partial E}{\partial u_{(s-1),p}} y_{(s-2),r} \; . \tag{3.26}$$

Definindo (analogamente ao resultado da Equação 3.20)

$$\delta_{(s-1),p} = -\frac{1}{2} \frac{\partial E}{\partial u_{(s-1),p}}, \tag{3.27}$$

pela regra da cadeia

$$\delta_{(s-1),p} = -\frac{1}{2} \frac{\partial E}{\partial y_{(s-1),p}} \frac{\partial y_{(s-1),p}}{\partial u_{(s-1),p}}, \tag{3.28}$$

onde $y_{(s-1),p}$ é a saída do neurônio p da camada $s-1$. Como $y_{(s-1),p} = \varphi(u_{(s-1),p}) \; ,$

$$\delta_{(s-1),p} = -\frac{1}{2} \frac{\partial E}{\partial y_{(s-1),p}} \frac{\partial \varphi(u_{(s-1),p})}{\partial u_{(s-1),p}}. \tag{3.29}$$

O fator $\dfrac{\partial E}{\partial y_{(s-1),p}}$ à direita da Equação (3.29) pode ser reformulado:

$$\frac{\partial E}{\partial y_{(s-1),p}} = \frac{\partial \left(\sum_{q=1}^{j} e_q^2 \right)}{\partial y_{(s-1),p}} = 2 \sum_{q=1}^{j} e_q \frac{\partial e_q}{\partial y_{(s-1),p}}. \tag{3.30}$$

Pela regra da cadeia,

$$\frac{\partial E}{\partial y_{(s-1),p}} = 2 \sum_{q=1}^{j} e_q \frac{\partial e_q}{\partial u_{s,q}} \frac{\partial u_{s,q}}{\partial y_{(s-1),p}}. \tag{3.31}$$

Os fatores $\dfrac{\partial e_q}{\partial u_{s,q}}$ e $\dfrac{\partial u_{s,q}}{\partial y_{(s-1),p}}$ à direita da Equação (3.31) também podem ser reformulados:

$$\frac{\partial e_q}{\partial u_{s,q}} = \frac{\partial (d_q - y_{s,q})}{\partial u_{s,q}} = -\frac{\partial y_{s,q}}{\partial u_{s,q}} = -\frac{\partial \varphi(u_{s,q})}{\partial u_{s,q}}, \tag{3.32}$$

$$\frac{\partial u_{s,q}}{\partial y_{(s-1),p}} = \frac{\partial \left(\sum_p y_{(s-1),p} \; w_{s,pq} \right)}{\partial y_{(s-1),p}} = w_{s,pq} \; . \tag{3.33}$$

Substituindo os resultados das Equações (3.32) e (3.33) na Equação (3.31),

$$\frac{\partial E}{\partial y_{(s-1),p}} = -2 \sum_{q=1}^{j} e_q \frac{\partial(\varphi(u_{s,q}))}{\partial u_{s,q}} w_{s,pq} \, . \tag{3.34}$$

Substituindo o resultado da Equação (3.34) na Equação (3.29) e rearranjando os fatores,

$$\delta_{(s-1),p} = \frac{\partial \varphi(u_{(s-1),p})}{\partial u_{(s-1),p}} \sum_{q=1}^{j} e_q \frac{\partial(\varphi(u_{s,q}))}{\partial u_{s,q}} w_{s,pq}. \tag{3.35}$$

De acordo com a definição de $\delta_{s,q}$ feita na Equação (3.16),

$$\delta_{(s-1),p} = \frac{\partial \varphi(u_{(s-1),p})}{\partial u_{(s-1),p}} \sum_{q=1}^{j} \delta_{s,q} \, w_{s,pq}. \tag{3.36}$$

Reformulando a Equação (3.27),

$$\frac{\partial E}{\partial u_{(s-1),p}} = -2\delta_{(s-1),p} \, . \tag{3.37}$$

Deste modo,

$$\frac{\partial E}{\partial w_{(s-1),rp}} = -2\delta_{(s-1),p} \, y_{(s-2),r} \, . \tag{3.38}$$

Substituindo o resultado da Equação (3.37) na Equação (3.26),

$$\Delta w_{(s-1),rp} = 2\alpha \, \delta_{(s-1),p} \, y_{(s-2),r} \, . \tag{3.39}$$

Para redes com mais que duas camadas, as regras de atualização das camadas restantes podem ser obtidas, por indução, a partir das Equações (3.36) e (3.39). Para ajustar o peso sináptico $w_{(s-l),nr}$ (o peso sináptico do r-ésimo neurônio da camada $s-l$, $l \in \{2, 3, ...\}$, que pondera $y_{(s-l-1),n}$, a n-ésima saída da camada $s-l-1$), as equações de atualização são:

$$\delta_{(s-l),r} = \frac{\partial \varphi(u_{(s-l),r})}{\partial u_{(s-l),r}} \sum_{p} \delta_{(s-l+1),p} \, w_{(s-l+1),rp} \, , \tag{3.40}$$

$$\Delta w_{(s-l),nr} = 2\alpha \, \delta_{(s-l),r} \, y_{(s-l-1),n} \, . \tag{3.41}$$

3.5 Treinamentos Seqüencial e por Batelada

O algoritmo *backpropagation* descreve como utilizar um par de vetores (um vetor de entrada e o respectivo vetor-objetivo) para calcular ajustes que produzam

a redução de uma medida de custo E (função do vetor de entrada utilizado, do respectivo vetor-objetivo e da configuração de pesos da rede). Porém, o treinamento de uma rede *feed-forward* normalmente é realizado através de sucessivas aplicações do algoritmo *backpropagation* sobre diversos pares de vetores pertencentes a um conjunto de treinamento (composto por N vetores de entrada e pelos respectivos vetores-objetivo). O treinamento visa a minimizar o custo médio $\bar{E} = \frac{1}{N}\sum_{i=1}^{N} E[i]$, onde $E[i]$ é o custo associado ao i-ésimo par de vetores utilizados durante o treinamento.

Existem dois métodos principais para aplicar o algoritmo *backpropagation* sobre o conjunto de vetores de treinamento: o método de treinamento seqüencial e o de treinamento por batelada.

No treinamento seqüencial, dentro de cada época (período em que são apresentados os pares de vetores do conjunto de treinamento, uma vez cada, até todos serem utilizados), os ajustes dos pesos sinápticos são realizados a cada iteração do algoritmo *backpropagation*. Inicialmente, um par de vetores do conjunto de treinamento é selecionado e uma iteração do algoritmo é aplicada. Nesta iteração são realizados tanto os cálculos dos ajustes, de acordo com as Equações (3.22) e (3.39), quanto as atualizações dos pesos sinápticos. O procedimento é então repetido seqüencialmente para os outros pares de vetores do conjunto de treinamento até que todos sejam utilizados. Após cada época, se um critério de parada pré-estabelecido não for atendido, é necessário iniciar uma nova época de treinamento. Neste método, as atualizações realizadas em uma iteração podem reduzir o custo para o padrão apresentado, mas também aumentar o custo médio do conjunto de treinamento. Para um número elevado de iterações, porém, o custo médio decresce [29].

No treinamento por batelada os pesos sinápticos são atualizados somente após cada época. Inicialmente um par de vetores do conjunto de treinamento é selecionado e uma iteração do algoritmo *backpropagation* é aplicada. Nesta iteração são realizados apenas os cálculos dos ajustes, de acordo com as Equações (3.22) e (3.39). Os valores dos ajustes são, então, armazenados. O procedimento é repetido para os outros pares de vetores do conjunto de treinamento até que todos sejam utilizados. Cada peso sináptico é, então, atualizado com a média de seus N ajustes, calculados dentro da época. Assim como no treinamento seqüencial, será necessário

iniciar uma nova época de treinamento se um critério de parada pré-estabelecido não for atendido após a atualização dos pesos sinápticos.

O critério de parada utilizado nesta tese envolve a avaliação da função de custo $\bar{E} = \sum_{i=1}^{M} E[i]$, aplicada a um conjunto de validação formado por M vetores de entrada (diferentes dos vetores de entrada do conjunto de treinamento, porém pertencentes às mesmas classes) e seus respectivos vetores-objetivo. Após o fim das atualizações realizadas a cada época, o custo E é avaliado para os pares de vetores do conjunto de validação, processados na configuração corrente da rede. Cada nova configuração da rede que gerar um custo inferior ao menor custo anteriormente avaliado é armazenada, e uma nova época de treinamento é iniciada. O critério de parada só é atendido quando o custo para o conjunto de validação aumentar consistentemente durante um número mínimo de épocas. A configuração final da rede (aquela que gerar o menor custo para o conjunto de validação dentre as configurações testadas) é utilizada para classificar novos sinais. Este método é utilizado para evitar o excesso de treinamento. O aumento do custo para o conjunto de validação indica a redução do desempenho da rede na classificação de vetores que não fazem parte do conjunto de treinamento, mas que pertencem às classes treinadas.

A avaliação do desempenho da rede na classificação de vetores inéditos deve ser feita sobre um conjunto de teste formado por vetores diferentes dos vetores dos conjuntos de treinamento e de validação, porém pertencentes às mesmas classes.

3.6 Algoritmo Rprop

RIEDMILLER e BRAUN [34] desenvolveram o algoritmo Rprop (ou *Resilient Backpropagation*) para treinamento por batelada como uma alternativa para o treinamento de redes *feed-forward* capaz de evitar falhas de convergência que podem ocorrer quando o algoritmo *backpropagation* é utilizado. Pela Equação (3.9), repetida abaixo para facilitar a leitura, a evolução de um treinamento realizado com o algoritmo *backpropagation* depende tanto do valor de α quanto do comportamento da derivada parcial $\frac{\partial E}{\partial w_{c,pq}}$:

$$\Delta w_{c,pq} = -\alpha \frac{\partial E}{\partial w_{c,pq}}. \tag{3.42}$$

A escolha de um valor apropriado para a taxa α pode não ser suficiente para garantir a convergência do processo de treinamento porque a evolução da magnitude de $\dfrac{\partial E}{\partial w_{c,pq}}$ é imprevisível. Alternativas de treinamento inspiradas no algoritmo *backpropagation* que utilizam termos de momento ou taxas de aprendizado adaptativas [35, 36] também são suscetíveis a este problema, porém em diferentes escalas.

No algoritmo Rprop, o valor do ajuste $\Delta w_{c,pq}$ não é proporcional à magnitude de $\dfrac{\partial E}{\partial w_{c,pq}}$. Em vez disto, o ajuste é definido pela evolução do sinal de $\dfrac{\partial E}{\partial w_{c,pq}}$ (obtido, para redes de duas camadas, através das Equações (3.21) e (3.38)) de acordo com a seguinte heurística:

Sendo $\dfrac{\partial \tilde{E}[t]}{\partial w_{c,pq}}$ igual a soma das derivadas parciais da função de custo E em relação ao peso sináptico $w_{c,pq}$ para todos os pares de vetores de treinamento apresentados em uma época t, dado um valor adaptativo $\Delta_{c,pq}$ referente ao peso sináptico $w_{c,pq}$,

$$\Delta_{c,pq}[t] = \begin{cases} \alpha^{+}\Delta_{c,pq}[t-1], & \text{se } \dfrac{\partial \tilde{E}[t-1]}{\partial w_{c,pq}} \dfrac{\partial \tilde{E}[t]}{\partial w_{c,pq}} > 0 \\ \alpha^{-}\Delta_{c,pq}[t-1], & \text{se } \dfrac{\partial \tilde{E}[t-1]}{\partial w_{c,pq}} \dfrac{\partial \tilde{E}[t]}{\partial w_{c,pq}} < 0 \\ \Delta_{c,pq}[t-1], & \text{se } \dfrac{\partial \tilde{E}[t-1]}{\partial w_{c,pq}} \dfrac{\partial \tilde{E}[t]}{\partial w_{c,pq}} = 0, \end{cases}$$

onde $0 < \alpha^{-} < 1 < \alpha^{+}$, e

$$\Delta w_{c,pq}[t] = \begin{cases} -\Delta_{c,pq}[t], & \text{se } \dfrac{\partial \tilde{E}[t]}{\partial w_{c,pq}} > 0 \\ +\Delta_{c,pq}[t], & \text{se } \dfrac{\partial \tilde{E}[t]}{\partial w_{c,pq}} < 0 \\ 0, & \text{se } \dfrac{\partial \tilde{E}[t]}{\partial w_{c,pq}} = 0. \end{cases}$$

Assim, o valor de $\Delta_{c,pq}$ aumenta enquanto a derivada parcial $\dfrac{\partial \tilde{E}}{\partial w_{c,pq}}$ mantiver o mesmo sinal (incrementando a velocidade de treinamento). Se a derivada trocar de sinal (na possível passagem por um mínimo local), o valor de $\Delta_{c,pq}$ diminui. Para a época t, se a derivada for positiva, o ajuste $\Delta w_{c,pq}[t]$ recebe o negativo do valor $\Delta_{c,pq}$ (buscando corrigir o sentido do treinamento). Se a derivada for negativa, o sentido não precisa ser corrigido, e $\Delta w_{c,pq}[t]$ recebe o valor $\Delta_{c,pq}$. Além destas regras,

quando ocorrer troca do sinal da derivada na passagem da época $t-1$ para a época t, o ajuste realizado na época $t-1$ deve ser revertido ($\Delta w_{c,pq}[t] = -\Delta w_{c,pq}[t-1]$) e o novo valor, reduzido, de $\Delta_{c,pq}$, calculado na época t, deve ser utilizado na época seguinte ($\Delta_{c,pq}[t+1] = \Delta_{c,pq}[t]$).

Capítulo 4

Metodologia para Identificação de Notas de Violão

4.1 Introdução

Os métodos para identificação de notas desenvolvidos nesta tese podem ser separados em dois grupos principais. No primeiro grupo estão os métodos que utilizam apenas uma rede neural e uma representação espectral na identificação das notas de cada combinação. No segundo grupo estão os métodos que utilizam duas redes neurais e duas representações espectrais na identificação das notas de cada combinação. Os métodos do segundo grupo têm uma rede para identificar a nota mais grave de cada combinação e outra para encontrar os intervalos entre a nota mais grave e as notas restantes. A segunda rede recebe como vetor de entrada uma versão 'transposta' da representação espectral. As duas redes são utilizadas em seqüência. Após conhecer a estimativa para a nota mais grave, resultado do processamento com a primeira rede, o espectro da CQT é alterado (de modo similar ao descrito na Seção 2.3) para que a componente analisada sobre o *pitch* da nota mais grave se torne a primeira componente do espectro. O novo espectro é então analisado com a segunda rede para obter estimativas dos intervalos entre a nota mais grave da combinação e as notas restantes.

4.2 Banco de Dados

Para treinar e testar as redes neurais desenvolvidas nesta tese foram utilizados bancos de dados compostos por registros de notas individuais de violão e piano. Os registros de notas de violão foram obtidos do banco de gravações de áudio *RWC Music Database: Musical Instrument Sound Database* [37] e de um banco criado por mim, referido aqui como banco RNV (Registros de Notas de Violão). Ambos os bancos são compostos por registros digitais monaurais com resolução de 16 bits e taxa de amostragem de 44100 Hz.

As gravações do banco RWC[1] utilizadas nesta tese, nomeadas com as siglas 091CGAFP, 091CGAFM, 091CGAFF, 092CGAFP, 092CGAFM, 092CGAFF, 093CGAFP, 093CGAFM e 093CGAFF, contêm seqüências de notas de 3 violões diferentes (designados por '091CG', '092CG' e '093CG'). Estas gravações foram realizadas com a técnica *apoyando*[2] (indicada pela letra 'A'), sem palheta (indicado pela penúltima letra de cada sigla, 'F'), executadas com 3 níveis diferentes de dinâmica (indicados pela última letra de cada sigla, 'P' para *piano*, 'M' para *mezzo* e 'F' para *forte*). Cada gravação é composta por uma seqüência de 78 sons de notas individuais, registros das 13 notas mais graves de cada corda. Pode-se notar que a variação de amplitude entre notas iguais em gravações com dinâmicas diferentes é, muitas vezes, pequena. Isto pode ter sido causado por variações na amplificação dos sinais ou na distância do microfone ao instrumento. A Figura 4.1 contém a representação de um braço de violão com o desenho das cordas apenas sobre as posições utilizadas para gerar estes registros, incluindo 6 posições sobre o capotraste (a parte do braço que suporta as cordas e que serve como nó quando as cordas vibram livremente). Como o som de uma mesma nota pode ser obtido utilizando cordas diferentes, com algumas exceções (como visto na Seção 1.7), cada nota pode ter um, dois ou três registros em cada gravação. Por exemplo: a nota Lá 2 tem 2 registros: um obtido tocando a 6^a corda (mantida pressionada sobre a 5^a casa a partir do capotraste[3]) e outro obtido tocando a 5^a corda solta. A nota Ré 4 tem 3 registros

[1] A partir deste ponto, por brevidade, o banco *RWC Music Database: Musical Instrument Sound Database* será referido apenas como banco RWC.

[2] Na técnica *apoyando* o músico apoia, após o plectro, o dedo ou palheta na corda adjacente.

[3] As casas são os espaços entre trastes.

em cada gravação: o primeiro obtido tocando a 4^a corda (mantida pressionada sobre a 12^a casa), o segundo obtido tocando a 3^a corda (mantida pressionada sobre a 7^a casa) e o terceiro obtido tocando a 2^a corda (mantida pressionada sobre a 3^a casa). Cada gravação contém de um a três registros de 37 notas diferentes.

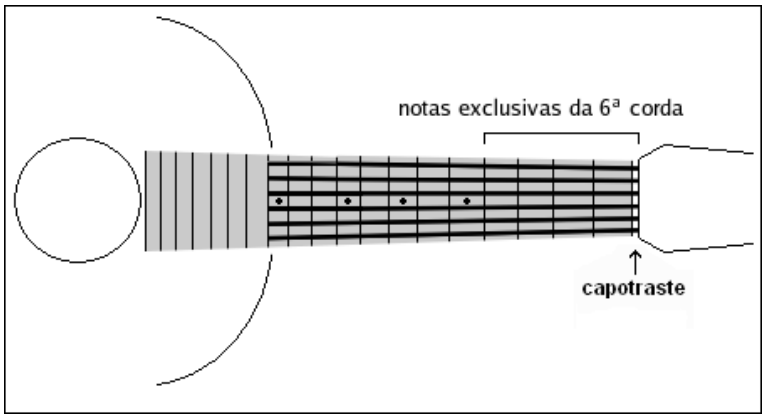

Figura 4.1: Representação do braço de um violão com as cordas desenhadas somente sobre a região utilizada para realizar as gravações da base RWC.

Para a construção do banco RNV foram gravados sons de 5 violões diferentes, nomeados A, B, C, D e E. A Figura 4.2 contém a representação de um braço de violão com o desenho das cordas sobre as 78 posições utilizadas para gerar estes registros, novamente incluindo as 6 posições sobre o capotraste. Com esta escolha foram gravadas todas as 44 notas diferentes que podem ser obtidas com violão normal, com dois registros de cada uma das 34 notas que podem ser tocadas em cordas diferentes. As 10 notas restantes têm apenas um registro.

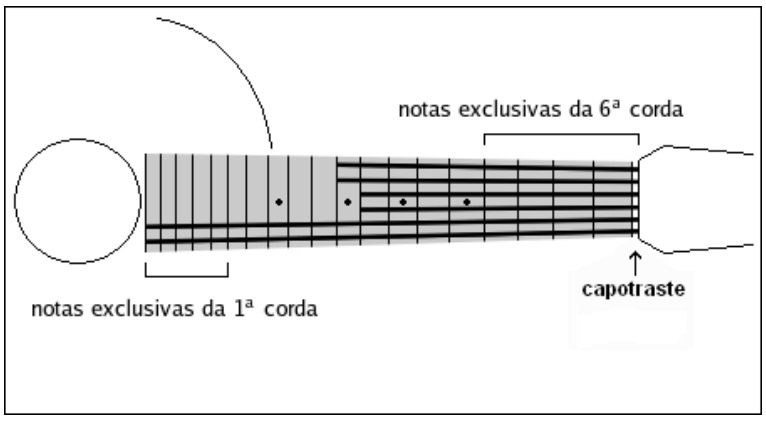

Figura 4.2: Representação do braço de um violão com as cordas desenhadas somente sobre a região utilizada para realizar as gravações do banco RNV.

Durante as gravações, apenas a corda utilizada na geração da nota podia vi-

brar. As 5 demais cordas permaneciam abafadas. Todas as notas foram executadas com dinâmica *mezzo*. As gravações foram realizadas em um ambiente sem tratamento acústico mas silencioso. Foi utilizado um microfone Shure®, modelo SM81. Nas gravações de notas obtidas a partir de posições entre o capotraste e a 12^a casa, o microfone foi voltado para a 12^a casa do braço. Nas gravações de notas obtidas a partir de posições além da 12^a casa, o microfone foi apontado para a região entre a casa pressionada e a cavidade do tampo. A distância entre o microfone e o braço foi mantida em torno de 5 cm, com pequenas variações para evitar a saturação da cápsula.

4.3 Segmentação

Para utilizar as gravações descritas na Seção 4.2 no desenvolvimento das redes foi necessário segmentá-las, separando cada uma das notas, a partir de seu *onset*, em um arquivo independente.

A segmentação de cada registro do banco RNV foi realizada através da análise visual das formas de onda de cada gravação. Este método é comumente chamado de 'método manual', porque não faz uso de algoritmos computacionais para detecção de *onsets*. Cada registro foi disponibilizado no banco de dados como uma gravação independente, já segmentada.

Uma parte dos *onsets* das gravações da base RWC foi especificada de acordo com marcações disponibilizadas por YEH [38]. Outra parte foi especificada utilizando um algoritmo de detecção de *onsets*, em desenvolvimento, gentilmente cedido por Jorge Costa Pires Filho. Um grupo de marcações foi realizada manualmente, assim como no banco RNV. Todas as marcações foram revisadas, e corrigidas quando necessário, inclusive as marcações disponibilizadas por Chunghsin Yeh. O final de cada segmento, o *offset*, foi estabelecido como o instante da amostra anterior ao *onset* da nota seguinte ou, no caso do último registro de cada gravação, como o instante da última amostra da gravação. As marcações de *onsets* da base RWC utilizadas nesta teste estão listadas no Apêndice A.1.

4.4 Criação dos *Kernels* da CQT

Como a quantidade de amostras $N[k_{\text{cq}}]$ utilizadas no cálculo da CQT cresce à medida que a análise se estende para freqüências mais baixas (Equação (2.5)), intervalos utilizados na análise de gravações de violão podem compreender trechos não-estacionários de sinal. Reduzir a seletividade da CQT diminui a quantidade de amostras necessárias para calcular cada componente da transformada, porém uma análise com baixa seletividade pode não possibilitar a distinção entre parciais presentes no sinal. Para calcular uma componente da CQT sobre o *pitch* de Mi 2, a nota mais grave do violão, com seletividade $Q \approx 68{,}75$ correspondente a uma resolução de 1/4 de semitom, é necessário utilizar um intervalo de 0,83 ms.

BROWN [28] propôs alterar a seletividade dos *kernels* da CQT em função da faixa do espectro sob análise para obter uma representação que tenha boa resolução nas regiões mais elevadas do espectro (que podem concentrar superposições de parciais), sem que seja necessário utilizar um grande número de amostras para a análise de componentes nas regiões mais baixas da faixa do instrumento (onde é menor o efeito de superposição de parciais). Nesta tese foram utilizados quatro valores diferentes para a seletividade ao longo da representação espectral.

Para o primeiro e para o segundo grupo de métodos foram criados *kernels* da CQT para a análise de componentes a partir do *pitch* da nota mais grave do violão (Mi 2; *pitch* $\approx 82{,}41$ Hz) até aproximadamente 5274,04 Hz. Com esta escolha para o limite superior é possível representar cinco parciais da nota mais aguda do instrumento (Si 5; *pitch* $\approx 987{,}77$ Hz) e, se os efeitos de inarmonicidade não forem muito elevados, cinco parciais da nota seguinte (Dó 6; *pitch* $\approx 1046{,}50$ Hz).

Os *kernels* da 1^a oitava da transformada foram criados com resolução freqüencial $q = 2^{\frac{1}{12}}$, correspondente a 1 semitom. Deste modo, a duração do intervalo necessário para o cálculo da componente sobre o *pitch* de Mi 2 é aproximadamente igual a 0,20 s. Foram criados dois grupos de *kernels* para a análise da 2^a oitava: o 1^o com resolução freqüencial $q = 2^{\frac{1}{24}}$, correspondente a 1/4 de tom e o 2^o com resolução freqüencial $q = 2^{\frac{1}{36}}$, correspondente a 1/6 de tom. O 1^o grupo abrange, com 14 componentes, as 7 primeiras notas dessa oitava. O 2^o grupo abrange, com 15 componentes, as 5 últimas notas dessa oitava. A troca de seletividade dentro da oitava é justificada pelo possível aumento no número de parciais, dependendo de

quais notas são analisadas, presentes dentro da oitava. A 3^a parcial referente à nota mais grave do violão ocorre em aproximadamente 247,22 Hz (assumindo que não existe inarmonicidade), enquanto a 1^a componente do 2^o grupo é analisada sobre a freqüência 246,94 Hz. A partir da 3^a oitava os *kernels* da transformada foram realizados com resolução freqüencial $q = 2^{\frac{1}{48}}$, correspondente a 1/8 de tom. No total, uma representação freqüencial criada utilizando estes *kernels* contém 234 componentes: 12 na 1^a oitava, 29 na 2^a oitava e 193 a partir da 1^a componente da 3^a oitava até a última componente da transformada.

Para o segundo grupo de métodos são utilizados, além dos *kernels* descritos anteriormente, *kernels* complementares para o cálculo dos espectros transpostos. A rede utilizada para estimar os intervalos entre a nota mais grave de cada combinação e as notas restantes processa representações espectrais que têm a mesma seqüência de seletividade por componente descrita para o primeiro grupo de métodos, porém as transformadas devem ser calculadas a partir do *pitch* da nota mais grave da combinação. Assim, se a nota mais grave for Ré 3 ($pitch \approx 146{,}83$ Hz), a CQT será calculada, utilizando 234 *kernels*, com componentes a partir de aproximadamente 146,83 Hz. Deste modo, transformadas com a primeira componente a partir do *pitch* da nota Fá 4 ($pitch \approx 349{,}23$ Hz) teriam componentes calculadas acima da freqüência de Nyquist [39]. Os *kernels* referentes a estas componentes não são calculados, e os valores de suas componentes são preenchidos com zeros durante o cálculo das transformadas. Como a seletividade das componentes varia em função da nota mais grave da combinação, só é possível aproveitar parte das componentes calculadas no processo de identificação da nota mais grave. Apenas as componentes que tiverem a seletividade mantida podem ser usadas para gerar a nova representação espectral (Seção 2.3).

4.5 Criação das Combinações de Notas Musicais

Os bancos de dados descritos na Seção 4.2 foram utilizados para criar sons formados por combinações de diferentes notas musicais. Representações espectrais destes sons servem como vetores de entrada para as redes neurais. Cada vetor de entrada foi associado a um vetor-objetivo de 44 elementos que indica quais notas estão

presentes em cada combinação, possibilitando realizar o treinamento supervisionado das redes.

As combinações de notas foram realizadas computacionalmente, criando sons com diferentes graus de polifonia, com até seis notas simultâneas. Usar combinações realizadas computacionalmente, em vez de gravações de um músico, possibilita criar uma grande quantidade de exemplos para o treinamento das redes sem a necessidade de gravar individualmente cada combinacão de notas. Por outro lado, os sons gerados com este procedimento não apresentam efeitos de acoplamento entre modos de vibração de cordas diferentes, que podem ocorrer durante a execução do instrumento. Espera-se que estes efeitos possam ser desconsiderados durante o desenvolvimento das redes sem causar impacto significativo na aplicação prática do método.

Nenhuma limitação com base em regras harmônicas foi utilizada na escolha das combinações de notas utilizadas. Na música ocidental existem várias regras de construção harmônica – regras para a combinação musical de notas simultâneas – e muitos músicos que ignoram, propositalmente ou não, estas regras. Deste modo, desenvolver um sistema de transcrição estabelecendo quais combinações são válidas e quais não são implicaria a polarização dos resultados para um determinado conjunto de regras e não refletiria a riqueza de possibilidades que podem ser encontradas em composições da música ocidental moderna. Estabelecer, porém, algum tipo de limitação harmônica pode ser útil para obter transcrições voltadas para determinados estilos musicais ou como um passo intermediário para a criação de um sistema mais abrangente.

As notas presentes em um instante qualquer de uma gravação real podem estar em etapas diferentes na evolução de suas envoltórias. Algumas notas podem, por exemplo, estar no período de ataque, enquanto outras estão no período de decaimento ou de sustentação. As notas também podem ter dinâmicas diferentes.

Para construir um conjunto realista de sinais, com eventos similares aos que podem ocorrer em gravações musicais, alguns cuidados foram tomados na criação das combinações:

a. Para cada instrumento do banco de dados foi gerada uma série independente de combinações dos registros disponíveis. O primeiro registro era escolhido aleatoria-

mente, de uma distribuição uniforme, entre os 78 registros disponíveis. Em seguida, todos os registros de notas tocadas na mesma corda eram excluídos da escolha seguinte. Também eram excluídos todos os registros da mesma nota tocados em outras cordas. O segundo registro era escolhido aleatoriamente, de uma nova distribuição uniforme, entre os registros restantes. Novamente todos os registros de notas tocadas na mesma corda e registros da mesma nota tocados em outras cordas eram excluídos da escolha seguinte. Este processo era repetido até se completar o grau de polifonia desejado. Para grau de polifonia igual a dois, foram esgotadas todas as combinações possíveis segundo esta regra (2480 combinações para cada violão da base RWC e 2427 combinações para cada violão da base RNV). Para graus de polifonia maiores que dois, foram escolhidas de 2750 até 4000 combinações diferentes para cada violão disponível, de acordo com o experimento realizado.

Na prática, o número de casas sobre o braço do instrumento que podem ser alcançadas simultaneamente é limitado pelo alcance dos dedos do músico. Assim, foram simuladas combinações de posições que não são possíveis no instrumento. Porém, o projeto do violão permite ao músico tocar muitas dessas combinações a partir de outras posições. Por exemplo, não é possível pressionar simultaneamente a 1^a casa da 6^a corda (Fá 2) e a 10^a casa da 2^a corda (Lá 4), mas é possível tocar as mesmas notas desta combinação, simultaneamente, pressionando a 1^a casa da 6^a corda (Fá 2) e a 5^a casa da 1^a corda (Lá 4).

b. Cada combinação era formada por trechos de registros segmentados de acordo com os objetivos apresentados na Seção 1.10. Os trechos que deveriam compreender aproximadamente o período de ataque e decaimento (como em combinações determinadas no 3^o e no 4^o objetivo) eram segmentados a partir da primeira amostra dos registros. Os trechos que deveriam compreender aproximadamente o período de sustentação (como em combinações determinadas em todos os objetivos) eram segmentados a partir da amostra 10001 dos registros. Os trechos que deveriam compreender aproximadamente o período de liberação (como em combinações determinadas no 3^o e no 4^o objetivo) eram segmentados a partir da amostra 20001 dos registros. Todos os segmentos tiveram a duração do maior intervalo necessário para o cálculo da CQT, aproximadamente 0,20 s. A escolha das amostras 10001 e 20001 como posições associadas aos inícios dos períodos de sustentação e liberação

foi realizada empiricamente, buscando valores coerentes com os inícios destes períodos na maioria dos registros utilizados. A detecção automática dos períodos do modelo ADSR [40] deve ser estudada em trabalhos futuros.

c. Antes de compor cada combinação, as dinâmicas dos trechos utilizados eram escolhidas, de acordo com os objetivos apresentados na Seção 1.10, entre *forte*, *mezzo* e *piano*. Todos os trechos segmentados das bases RWC e RNV foram normalizados pela norma quadrática e em seguida, de acordo com a dinâmica escolhida, poderiam ter suas amplitudes alteradas. Quando a dinâmica escolhida era *forte* (como em combinações determinadas no 2^o e no 3^o objetivo), a amplitude era mantida. Quando a dinâmica escolhida era *mezzo* (como em combinações determinadas em todos os objetivos), a amplitude era alterada, formando sinais com $-10\,\mathrm{dB}$ de potência em relação aos segmentos normalizados. Quando a dinâmica escolhida era *piano* (como em parte das combinações criadas para o 4^o objetivo), a amplitude era alterada, formando sinais com $-20\,\mathrm{dB}$ de potência em relação aos segmentos normalizados. Os registros da base RWC, que possui gravações de notas executadas com dinâmicas *forte*, *mezzo* e *piano*, eram selecionados de acordo com as dinâmicas escolhidas. Como não existem variações dinâmicas na base RNV, diferentes potências eram associadas a qualquer registro desta base.

d. Sinais de notas simples também foram utilizados nos treinamentos e testes das redes, do mesmo modo que sinais polifônicos. Cada registro disponível, incluindo as diferentes versões de dinâmica dos registros da base RWC, foi segmentado em trechos de aproximadamente 0,20 s, com inícios a partir da primeira amostra, da amostra 10001 e da amostra 20001, de acordo com os objetivos apresentados na Seção 1.10.

As combinações foram criadas através da soma dos vetores compostos pelos elementos de cada segmento. Após a soma, cada combinação foi normalizada por sua norma quadrática.

4.6 Treinamento das Redes Neurais

Para cada combinação de notas foram calculadas duas transformadas através do algoritmo rápido da CQT (Seção 2.2). A primeira transformada, para aplicação

no primeiro e no segundo grupo de métodos, foi calculada com componentes a partir do *pitch* da nota Mi 2. A segunda transformada, para aplicação apenas no segundo grupo de métodos, foi calculada com componentes a partir do *pitch* da nota mais grave de cada combinação.

Na maioria dos testes, os vetores de entrada das redes neurais foram formados pelos valores absolutos das componentes de cada transformada. Os vetores-objetivo foram formados com 44 elementos, cada um correspondente a uma das notas do violão. A presença de cada nota foi indicada pelo valor 1 no elemento correspondente. As notas ausentes foram indicadas pelo valor 0. As dinâmicas e amostras iniciais escolhidas para cada combinação foram armazenadas para uso na análise dos resultados.

Os pares formados pelos vetores de entrada e vetores-objetivo foram divididos em três conjuntos: um de treino, um de teste e um de validação. O conjunto de treino continha os pares formados a partir de combinações das notas dos violões A, C e 091CG. O conjunto de validação continha os pares formados a partir de combinações das notas dos violões D, E e 093CG. O conjunto de teste continha os pares formados a partir de combinações das notas dos violões B e 092CG.

No desenvolvimento das redes utilizadas nas análises referentes ao **1º objetivo**, foram realizados experimentos com um número fixo de vetores nos grupos de treino e validação. Foram criados 3000 pares de vetores (entrada e objetivo) para cada instrumento, para cada grau de polifonia maior que dois. Foram criados 2480 pares de vetores referentes a combinações de duas notas para cada violão da base RWC e 2427 pares de vetores referentes a combinações de duas notas para cada violão da base RNV. Além destes, foram criados 78 pares de vetores referentes a notas simples para cada violão.

Para o grupo de teste foram criados 2750 pares de vetores para cada instrumento, para cada grau de polifonia maior que dois. Foram criados 2480 pares de vetores referentes a combinações de duas notas para cada violão da base RWC e 2427 pares de vetores referentes a combinações de duas notas para cada violão da base RNV. Além destes, foram criados 78 pares de vetores referentes a notas simples para cada violão.

No total foram criados 43568 pares de vetores para os conjuntos de treino

e validação e 27063 pares de vetores para o grupo de teste, todos construídos a partir de registros com dinâmica *mezzo*, extraídos aproximadamante do período de sustentação das notas.

No desenvolvimento das redes utilizadas nas análises referentes ao **2º objetivo**, os experimentos foram realizados com um número fixo de vetores nos grupos de treino e validação. Cada vetor foi construído a partir de registros com dinâmica *mezzo* (exceto por um com dinâmica *forte*), extraídos aproximadamante do período de sustentação das notas. Foram criados 3000 pares de vetores (entrada e objetivo) para cada instrumento, para cada grau de polifonia maior que dois. Foram criados 2480 pares de vetores referentes a combinações de duas notas para cada violão da base RWC e 2427 pares de vetores referentes a combinações de duas notas para cada violão da base RNV. Além destes, foram criados 78 pares de vetores referentes a notas simples para cada violão, todos com dinâmica *forte*.

Para o grupo de teste cada vetor foi construído a partir de registros com dinâmica *mezzo* (exceto por um com dinâmica *forte*), extraídos aproximadamante do período de sustentação das notas. Foram criados 2750 pares de vetores para cada instrumento, para cada grau de polifonia maior que dois. Foram criados 2480 pares de vetores referentes a combinações de duas notas para cada violão da base RWC e 2427 pares de vetores referentes a combinações de duas notas para cada violão da base RNV. Além destes, foram criados 78 pares de vetores referentes a notas simples para cada violão, todos com dinâmica *forte*.

No total foram criados 43568 pares de vetores para os conjuntos de treino e validação e 27063 pares de vetores para o grupo de teste.

No desenvolvimento das redes utilizadas nas análises referentes ao **3º objetivo**, os experimentos também foram realizados com um número fixo de vetores nos grupos de treino e validação. Cada vetor foi construído a partir de registros com dinâmica *mezzo* utilizando três possibilidades de segmentação: todos os segmentos extraídos aproximadamente do período que compreende o ataque e decaimento, todos os segmentos extraídos aproximadamente do período de sustentação e todos os segmentos extraídos aproximadamente do período de liberação.

Foram criados 1000 pares de vetores (entrada e objetivo) para cada violão, para cada grau de polifonia maior que dois, para cada segmentação possível. Foram

criados 826 pares de vetores referentes a combinações de duas notas para cada violão da base RWC, para cada segmentação possível e 809 pares de vetores referentes a combinações de duas notas para cada violão da base RNV, para cada segmentação possível. Além destes, foram criados 78 pares de vetores referentes a notas simples para cada violão, para cada segmentação possível.

Para o grupo de teste, cada vetor foi construído a partir de registros com dinâmica *mezzo* utilizando as mesmas três possibilidades de segmentação; porém, neste caso o grupo de teste foi dividido em 3 subgrupos, um para cada período aproximado da envoltória.

Foram criados 2750 pares de vetores para cada instrumento, para cada grau de polifonia maior que dois, para cada segmentação possível. Também foram criados 2480 pares de vetores referentes a combinações de duas notas para cada violão da base RWC, para cada segmentação possível e 2427 pares de vetores referentes a combinações de duas notas para cada violão da base RNV, para cada segmentação possível. Além destes, foram criados 78 pares de vetores referentes a notas simples para cada violão, para cada segmentação possível.

No total foram criados 44034 pares de vetores para os conjuntos de treino e validação contendo exemplos e 27063 pares de vetores para cada subgrupo de teste.

No desenvolvimento das redes utilizadas nas análises referentes ao **4º objetivo**, foram realizados experimentos variando o número de vetores nos conjuntos de treinamento e validação. Cada vetor foi construído a partir de registros com dinâmicas escolhidas aleatoriamente entre *forte*, *mezzo* e *piano*, extraídos de períodos escolhidos aleatoriamente entre ataque e decaimento, sustentação e liberação.

Para os conjuntos de treinamento e validação foram criados subconjuntos de 1000, 1500, 2000, 2500, 3000, 3500 e 4000 pares de vetores (entrada e objetivo) para cada instrumento, para cada grau de polifonia maior que dois. Foram criados 2480 pares de vetores referentes a combinações de duas notas para cada violão da base RWC e 2427 pares de vetores referentes a combinações de duas notas para cada violão da base RNV. Além destes, foram criados 702 pares de vetores referentes a notas simples para cada violão, abrangendo todas as combinações possíveis de 78 registros com 3 dinâmicas diferentes e 3 períodos de segmentação diferentes.

A quantidade de vetores utilizados no conjunto de teste foi mantida fixa para

permitir comparações entre resultados. Foram criados 2750 pares de vetores para cada instrumento, para cada grau de polifonia maior que dois. Foram criados 2480 pares de vetores referentes a combinações de duas notas para cada violão da base RWC e 2427 pares de vetores referentes a combinações de duas notas para cada violão da base RNV. Além destes, foram criados 702 pares de vetores referentes a notas simples para cada violão, abrangendo todas as combinações possíveis de 78 registros com 3 dinâmicas diferentes e 3 períodos de segmentação diferentes.

No total, para o 4º objetivo, foram realizados testes com conjuntos de treino e validação que continham 20504, 26504, 32504, 38504, 44504, 50504 e 56504 pares de vetores. O grupo de teste continha, sempre, 27843 pares de vetores.

Para todos os objetivos, os vetores de entrada foram escalonados para o uso com redes neurais. Os valores de cada componente foram reduzidos das médias de *ensemble* correspondentes (calculados apenas sobre o conjunto de treino), e divididos pelo dobro dos desvios-padrão de *ensemble* correspondentes (calculados apenas sobre o conjunto de treino). Apenas médias e desvios-padrão do conjunto de treino foram utilizados para evitar a polarização dos resultados em favor do grupo de teste ou do grupo de validação.

Os treinamentos foram realizados utilizando o algoritmo Rprop (Seção 3.6) e o critério de parada descrito na Seção 3.5. As constantes utilizadas no treinamento foram estabelecidas, empiricamente, como:

$$\alpha^+ = 1{,}05; \quad \alpha^- = 0{,}5; \quad \Delta w_{c,pq}[0] = 0{,}05.$$

Foi criado um limite máximo para $\Delta w_{c,pq}[n]$: $\Delta w_{\max} = 5$. Este limite foi proposto por RIEDMILLER [41] para evitar o aumento excessivo de valores de $\Delta w_{c,pq}[n]$.

Todas as redes desenvolvidas nesta tese têm duas camadas de neurônios com funções de ativação logística. Para encontrar topologias apropriadas para cada rede, a maioria dos testes envolveu variações na quantidade de neurônios na camada oculta.

Os pesos sinápticos de todas as redes foram inicializados com valores entre $-0{,}25$ e $0{,}25$, selecionados aleatoriamente dentro de uma distribuição uniforme. Esta faixa de valores foi escolhida para evitar a saturação de neurônios durante a inicialização das redes.

Capítulo 5

Implementação e Testes - Violão

5.1 Introdução

Neste capítulo são detalhados os métodos propostos para identificação de notas de violão e os resultados dos testes realizados.

Três medidas são apresentadas para avaliação dos resultados: o NER (*Note Error Rate*), o CER (*Chord Error Rate*)[1] e a acurácia. Estas medidas são realizadas sobre as classificações obtidas dos vetores do grupo de teste, sendo:
NC = total de notas classificadas corretamente nas combinações analisadas,
FN = total de falsos negativos (número de notas que deveriam ser classificadas como presentes nas combinações analisadas, mas não o foram),
FP = total de falsos positivos (número de notas que não deveriam ser classificadas como presentes nas combinações analisadas, mas o foram) e
N_{obj} = total de notas-objetivo associadas às combinações analisadas, (total de valores iguais a 1 no conjunto de vetores-objetivo).

- Primeira medida, NER:

  NER = somatório da quantidade falsos negativos e erros de inserção, dividido por N_{obj}. Os erros de inserção ocorrem quando a quantidade de notas classificadas como presentes em uma combinação excede o número de suas

[1] A palavra *chord* (acorde), no jargão musical, só é utilizada para combinações de três ou mais notas. Nesta dissertação, a medida CER também é utilizada para avaliar classificações de notas simples e de combinações de duas notas.

notas-objetivo. Para cada combinação, o erro de inserção é igual ao número de notas acusadas em excesso [18, 42].

- Segunda medida, CER:

  CER = total das combinações classificadas com pelo menos uma nota errada (erro falso positivo ou falso negativo), dividido pelo total de combinações testadas [42].

- Terceira medida:

$$\text{acurácia} = \frac{NC}{FN + FP + NC}.$$

  Como critério de avaliação de desempenho, a acurácia tem uma vantagem sobre o NER. A acurácia não pode exceder 100%, porém o NER pode (a soma de falsos negativos e erros de inserção pode exceder o número de notas-objetivo). O resultado da acurácia será no mínimo 0 (se todas as classificações estiverem erradas) e no máximo 1 (se todas as classificações estiverem corretas) [26]. Por isto, esta medida foi utilizada como critério para a escolha das redes desenvolvidas nesta tese. Uma medida similar à acurácia, chamada *score* [43], é utilizada na avaliação de transcrições consolidadas ao longo do tempo, nas quais o interesse não recai sobre as classificações de cada segmento de sinal analisado, mas sobre o resultado inferido da análise dinâmica dessas classificações. Este resultado normalmente é composto pelo *onset*, pela duração e pelo nome de cada nota estimada.

Todas as medidas descritas nesta seção são apresentadas em formato percentual nas próximas seções. Para comparação com resultados encontrados na literatura, algumas medidas são apresentadas em função do grau de polifonia dos segmentos analisados.

5.2 Métodos para Identificação de Notas de Violão - Objetivo 1

Nesta seção são detalhados os métodos desenvolvidos para identificar notas em combinações de registros com dinâmica *mezzo* a partir de segmentos extraídos

aproximadamente do período de sustentação de cada nota.

5.2.1 Métodos do Primeiro Grupo

Nos métodos do primeiro grupo é utilizada apenas uma rede neural para a análise das representações espectrais de cada combinação de notas.

5.2.1.1 Método 1A - Objetivo 1

No método 1A, os vetores de entrada da rede neural são formados pelos valores absolutos dos elementos das CQTs de cada combinação de notas. As CQTs são obtidas através do algoritmo rápido descrito na Seção 2.2, utilizando os *kernels* para métodos do primeiro grupo, descritos na Seção 4.4.

Cada nota é classificada como presente ou ausente de acordo com os valores dos elementos obtidos nos vetores de saída da rede treinada. As notas correspondentes aos elementos com valores maiores que 0,5 são classificadas como presentes. Se forem encontrados mais que 6 elementos com valores maiores que 0,5, apenas as 6 notas correspondentes aos 6 maiores elementos são classificadas como presentes na combinação correspondente. Se nenhum elemento tiver valor acima de 0,5, apenas a nota correspondente ao maior valor encontrado é classificada como presente.

O treinamento das redes foi realizado de acordo com a metodologia apresentada na Seção 4.6. Foram treinadas 3 redes diferentes, todas com 234 neurônios na camada oculta, o mesmo número de elementos do vetor de entrada. Cada realização foi inicializada com um grupo de pesos sinápticos diferentes, cada um deles, selecionado aleatoriamente de uma distribuição uniforme dos valores entre -0,25 e 0,25.

Tabela 5.1: Método 1A

rede	nº de épocas	acurácia %
1	41	68,6
2	38	**69,7**
3	40	68,8

Na Tabela 5.1 são mostrados os resultados da implementação do método 1A,

com as 3 realizações desenvolvidas, na classificação do conjunto de teste. O melhor resultado foi obtido na 2ª realização, com uma rede treinada em 38 épocas. Outros resultados desta classificação são mostrados nas Figuras 5.1 e 5.2, conjuntamente com resultados apresentados por BONNET e LEFEBVRE [18], obtidos através de seu método de identificação de notas em sinais polifônicos de violão. Para esta realização os resultados de NER e CER foram, respectivamente, 25,3% e 58,7%.

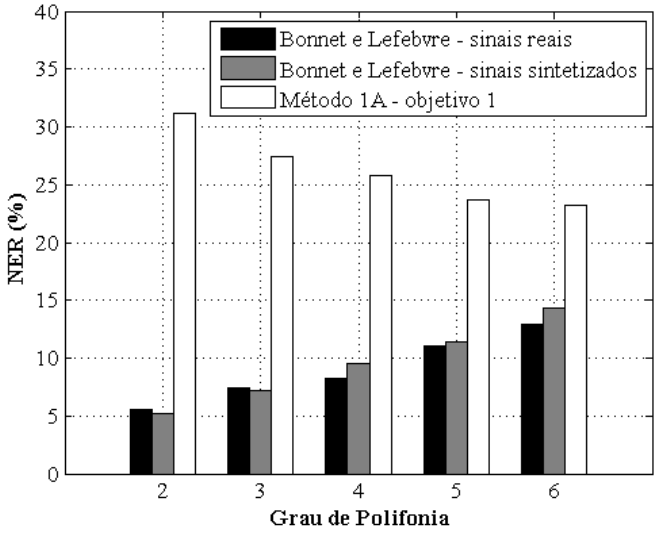

Figura 5.1: Percentuais do NER por grau de polifonia para o método de Bonnet e Lefebvre (na classificação de sinais reais e sintetizados de violão), e para o método 1A (na classificação do conjunto de teste).

Bonnet e Lefebvre realizaram suas análises sobre trechos de sinais correpondentes ao período de sustentação das notas (porém realizaram as segmentações de forma diferente da utilizada nesta tese). Os autores testaram seu método na classificação de dois conjuntos de sinais, um com sons sintetizados e outro com registros reais de acordes de violão[2]. Eles não apresentaram informações sobre a dinâmica das notas presentes nos sinais e não realizaram análises de sinais com notas simples.

Bonnet e Lefebvre apresentaram suas medições de erro em função do grau de polifonia dos acordes analisados. Eles realizaram medições do NER para ambos os conjuntos de sinais analisados e medições do CER apenas para o conjunto de registros reais de violão. Como o conjunto de teste desenvolvido para testar o

[2]Bonnet e Lefebvre se referem, também, a duas notas executadas simultaneamente como um acorde.

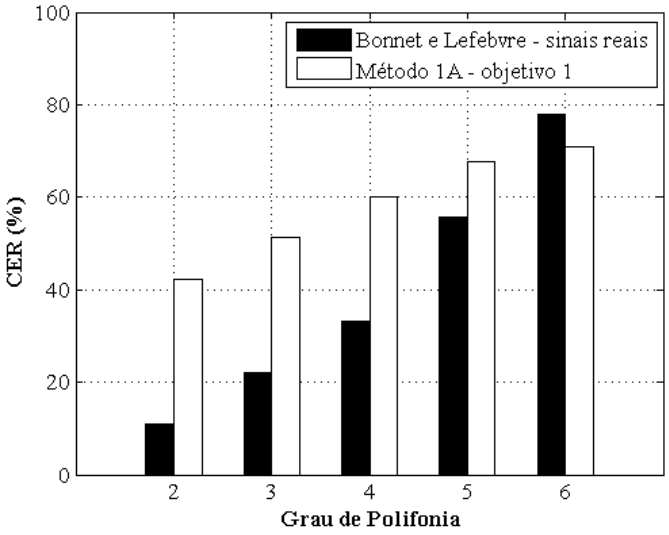

Figura 5.2: Percentuais do CER por grau de polifonia para o método de Bonnet e Lefebvre (na classificação de sinais reais de violão), e para o método 1A (na classificação do conjunto de teste).

método 1A é diferente dos conjuntos analisados por Bonnet e Lefebvre, a comparação entre os resultados apresentados nas Figuras 5.1 e 5.2 pode ser vista apenas como indicativa das vantagens de cada método.

Como pode ser observado na Figura 5.3, a maior parte dos erros de classificação obtidos utilizando o método 1A são do tipo falso positivo. Uma das dificuldades na identificação de notas é a falta de conhecimento prévio do grau de polifonia do segmento de sinal analisado. Alguns autores propuseram sistemas nos quais o grau de polifonia dos segmentos analisados é previamente conhecido [23, 24]. Acrescentar esta informação ao vetor de entrada e ao método de classificação pode favorecer os resultados, porque os erros de inserção (todos falsos positivos) são eliminados. Esta é a motivação para o desenvolvimento do Método 1B.

5.2.1.2 Método 1B - Objetivo 1

Os vetores de entrada utilizados no método 1B têm 6 elementos adicionais. Cada um representa um grau diferente de polifonia do violão. Um vetor de entrada associado a uma combinação de q notas é complementado com um vetor de 6 elementos, sendo o q-ésimo igual a 1 e os restantes iguais a zero. Dado o conhecimento prévio do grau de polifonia q, as notas referentes aos q maiores elementos do vetor

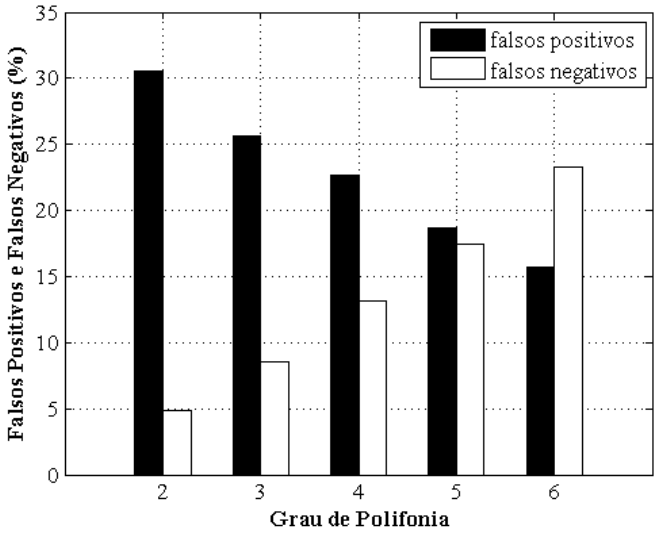

Figura 5.3: Totais de falsos positivos e falsos negativos por grau de polifonia para o método 1A.

de saída são classificadas como presentes na combinação.

As redes desenvolvidas para o método 1B têm, assim como as redes desenvolvidas para o método 1A, 234 neurônios na camada oculta. Na Tabela 5.2 são mostrados os resultados da implementação do método 1B, com as 3 realizações desenvolvidas, na classificação do conjunto de teste.

Tabela 5.2: Método 1B

rede	nº de épocas	acurácia %
1	41	69,2
2	44	69,9
3	40	**70,0**

O melhor resultado foi obtido na 3ª realização, com uma rede treinada em 40 épocas. Outros resultados desta classificação são mostrados nas Figuras 5.4 e 5.5, novamente em conjunto com os resultados obtidos por Bonnet e Lefebvre. Para esta realização os resultados de NER e CER foram, respectivamente, 17,7% e 45,9%.

Pode-se observar na Figura 5.6 que ocorreu redução no número de falsos positivos para graus de polifonia de 2 até 4. O aumento no número de falsos positivos para 6 notas simultâneas ocorreu por causa da escolha, obrigatória, de 6 notas

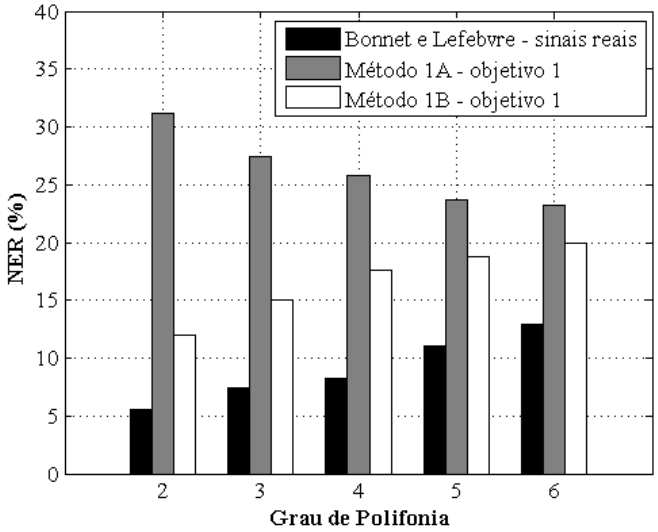

Figura 5.4: Percentuais do NER por grau de polifonia para o método de Bonnet e Lefebvre (na classificação de sinais reais de violão), para o método 1A (na classificação do conjunto de teste) e para o método 1B (na classificação do conjunto de teste).

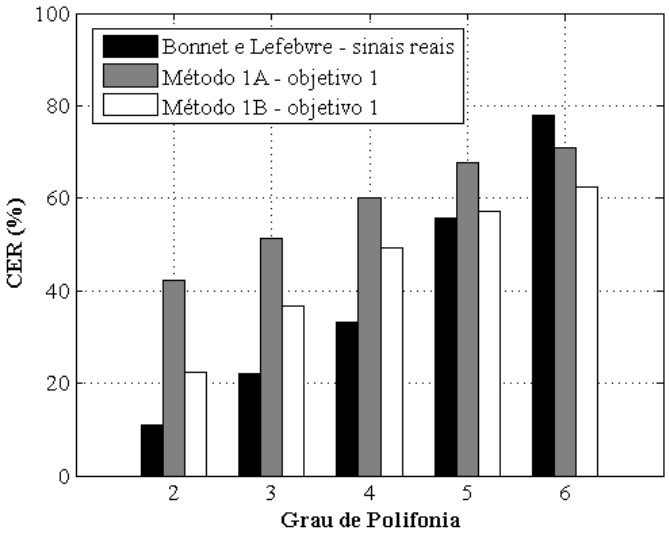

Figura 5.5: Percentuais do CER por grau de polifonia para o método de Bonnet e Lefebvre (na classificação de sinais reais de violão), para o método 1A (na classificação do conjunto de teste) e para o método 1B (na classificação do conjunto de teste).

para este grau de polifonia, estipulada no critério de classificação do método 1B. Utilizando este critério, elementos dos vetores de saída com valores abaixo de 0,5 também podem ser associados a notas (caso estejam entre os q maiores elementos). A ocorrência deste tipo de erro aumenta, neste caso, com o crescimento do grau de polifonia.

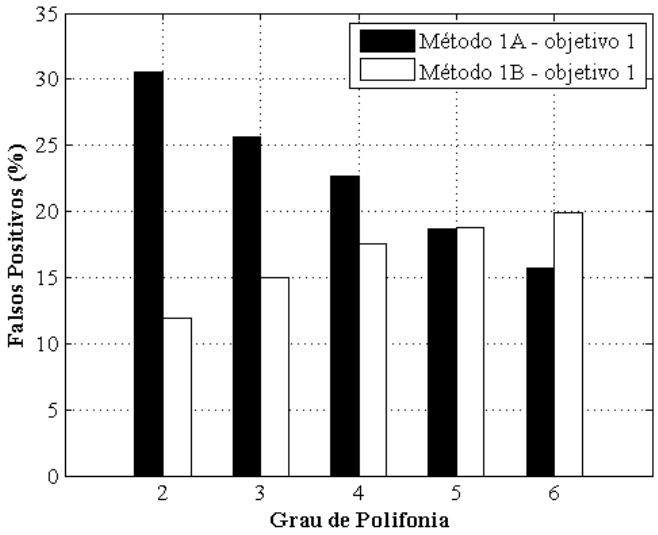

Figura 5.6: Falsos positivos por grau de polifonia para os métodos 1A e 1B.

Utilizar a informação do grau de polifonia para identificar as notas presentes em uma combinação melhora os resultados porque, principalmente, elimina os erros de inserção. Porém, esta informação pode não estar disponível. Para tentar melhorar os resultados, sem utilizar a informação do grau de polifonia, foram desenvolvidos os métodos do segundo grupo.

5.2.2 Métodos do Segundo Grupo

A identificação de notas musicais através da análise espectral é dificultada pelo grande número de combinações possíveis entre notas. Porém, dado o conhecimento de qual é a nota mais grave de cada combinação analisada, é possível transformar o problema de identificação de notas em um problema de identificação dos intervalos musicais entre a nota mais grave e as notas restantes. Utilizando a CQT, descrita na Seção 2.3, representações espectrais de notas diferentes – ou de combinações de notas diferentes, porém com os mesmos intervalos entre si – podem ter suas parciais alinhadas através de um deslocamento apropriado sobre a escala de

freqüências. Este procedimento cria um referencial comum para todos os vetores de entrada. Combinações com os mesmos intervalos são representadas por parciais alinhadas, mesmo se não tiverem as mesmas notas

Nos métodos do segundo grupo, a identificação é realizada utilizando duas redes em seqüência. Neles, a primeira rede é utilizada para identificar a nota mais grave de cada combinação e a segunda para encontrar os intervalos entre a nota mais grave de cada combinação e as notas restantes.

As representações espectrais utilizadas no treinamento das redes criadas para identificação da nota mais grave de cada combinação são as mesmas utilizadas para realizar os treinamentos do primeiro grupo de métodos.

As representações espectrais utilizadas no treinamento das redes criadas para identificação dos intervalos entre a nota mais grave e as notas restantes são realizadas utilizando os *kernels* criados para o segundo grupo de métodos, como descrito na Seção 4.4. As representações espectrais dos grupos de treino e validação são criadas, para cada combinação, a partir da componente sobre o *pitch* de sua nota mais grave. As representações espectrais do grupo de teste são criadas, para cada combinação, a partir da componente sobre o *pitch* da nota mais grave estimada.

Nas próximas seções são apresentados os métodos desenvolvidos para identificar a nota mais grave de cada combinação e, em seguida, o método desenvolvido para encontrar os intervalos entre a nota mais grave de cada combinação e as notas restantes.

5.2.2.1 Método 2A - 1ª etapa - Objetivo 1

No método 2A - 1ª etapa, as redes neurais recebem vetores de entrada formados apenas pelas representações espectrais de cada combinação. Cada vetor-objetivo é formado por 44 elementos, correspondentes, cada um, a uma nota diferente do violão. A presença da nota mais grave é indicada pelo valor 1 no elemento correspondente. Todas os outros elementos do vetor, inclusive os elementos correspondentes a outras notas presentes nas combinações analisadas, recebem o valor 0.

Para cada vetor de saída, a nota correspondente ao elemento com o maior valor é classificada como a nota mais grave da combinação.

A avaliação do desempenho é dada pelo percentual de combinações com erro

na classificação da nota mais grave (erro nmg). As estimativas das notas mais graves de cada combinação do conjunto de teste, obtidas com o método que gerar o menor erro nmg, são utilizadas para realizar as transposições dos vetores de entrada do conjunto.

Foram treinadas 3 redes diferentes, todas com 234 neurônios na camada oculta, o mesmo número de elementos do vetor de entrada. Na Tabela 5.3 são mostrados os resultados da implementação da 1ª estapa do método 2A com uma rede com 234 neurônios na camada oculta.

Tabela 5.3: Método 2A - 1ª etapa - Objetivo 1

rede	nº de épocas	erro nmg
1	36	**24,5**
2	41	25,1
3	48	24,8

O principal motivo para o elevado número de erros na identificação da nota mais grave, 24,5% para a melhor realização, pode ser inferido pela análise do histograma na Figura (5.7). Nele são dadas as quantidades de falsos positivos por nota da melhor realização da 1ª etapa do método 2A.

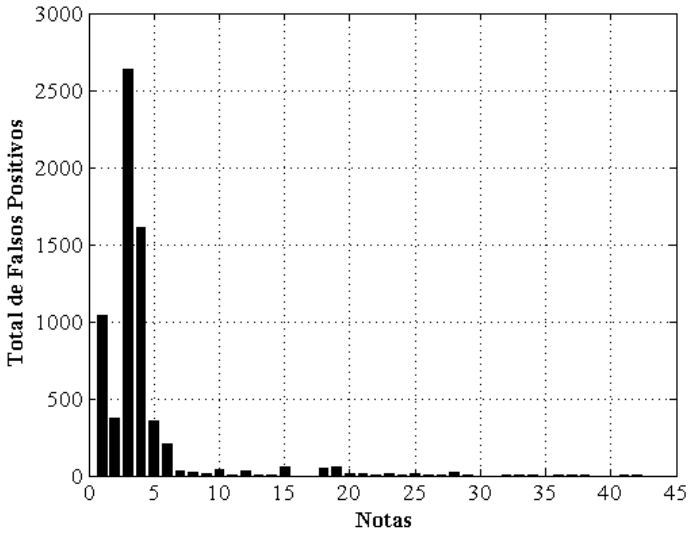

Figura 5.7: Histograma de Falsos Positivos.

Ocorreram muitos falsos positivos indicando notas da 1ª oitava do violão,

principalmente as notas 1, 3 e 4 (Mi 2, Fá♯2 e Sol 2). Estes erros foram, possivelmente, causados pela 1ª freqüência de ressonância do violão (Seção 1.7). A 1ª freqüência de ressonância é localizada tipicamente dentro da faixa entre 70 Hz e 140 Hz. Os *pitches* das notas Mi 2, Fá♯2 e Sol 2 são aproximadamente 82,41 Hz, 92,50 Hz e 98,00 Hz. Todos estão dentro desta faixa.

5.2.2.2 Método 2B - 1ª etapa - Objetivo 1

A 1ª etapa do método 2B foi criada visando à correção do problema de identificação da nota mais grave. Neste método, as 12 componentes da CQT calculadas sobre a 1ª oitava da faixa do instrumento são substituídas por 68 componentes de uma transformada discreta de Fourier sobre a mesma oitava.

Novamente, foram treinadas 3 redes diferentes para realizar a classificação, todas com 290 neurônios na camada oculta, quantidade igual ao número de elementos nos novos vetores de entrada. Na Tabela 5.4 são mostrados os resultados da implementação da 1ª etapa do método 2B.

Tabela 5.4: Método 2B - 1ª etapa - Objetivo 1

rede	nº de épocas	erro nmg
1	32	**23,1**
2	30	23,4
3	28	**23,1**

O desempenho deste método, apesar de melhor (23,1% nas duas melhores realizações), ainda é baixo. No histograma da Figura (5.8) são dadas as quantidades de falsos positivos por nota obtidos utilizando a primeira rede desenvolvida para este método. Novamente ocorreram muitos falsos positivos indicando notas da 1ª oitava do violão.

5.2.2.3 Método 2C - 1ª etapa - Objetivo 1

Para reduzir mais os erros, foi criado um método em que os vetores-objetivo eram iguais aos utilizados nos métodos do primeiro grupo, com a presença de cada

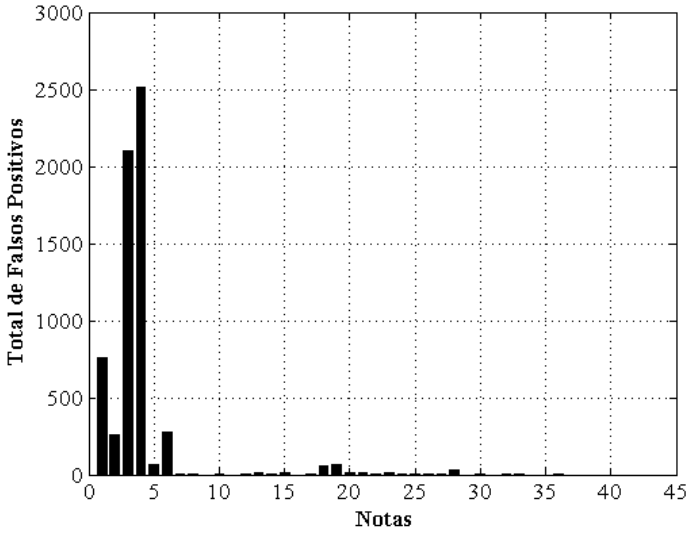

Figura 5.8: Histograma de Falsos Positivos por Notas.

nota (não só a da nota mais grave), indicada pelo valor 1 no elemento correspondente. Os vetores de entrada eram iguais aos criados para a 1ª etapa do método 2B.

Neste método, a classificação é realizada da mesma forma utilizada no método 1A. Cada nota é classificada como presente ou ausente de acordo com os valores dos elementos obtidos nos vetores de saída da rede treinada. As notas correspondentes aos elementos com valores maiores que 0,5 são classificadas como presentes. Se forem encontrados mais que 6 elementos com valores maiores que 0,5, apenas as 6 notas correspondentes aos 6 maiores elementos são classificadas como presentes na combinação correspondente. Se nenhum elemento tiver valor acima de 0,5, apenas a nota correspondente ao maior valor encontrado é classificada como presente.

Após este processo, a nota mais grave encontrada para cada combinação era selecionada como estimativa.

Foram treinadas 3 redes diferentes para realizar a classificação, todas com 290 neurônios na camada oculta. Na Tabela 5.5 são mostrados os resultados da implementação da 1ª etapa do método 2C.

Este método obteve o melhor desempenho entre os métodos da 1ª etapa (erro nmg = 21,9%).

Os métodos desenvolvidos para identificação da nota mais grave devem ser aperfeiçoados em trabalhos futuros.

Tabela 5.5: Método 2C - 1ª etapa - Objetivo 1

rede	nº de épocas	erro nmg
1	33	22,0
2	29	**21,9**
3	30	22,3

5.2.2.4 Método 2C - 2ª etapa - Objetivo 1

O desenvolvimento da 2ª etapa do método 2C independe do desenvolvimento dos métodos da 1ª etapa. Para testar seu desempenho foram realizados dois conjuntos de testes. No primeiro, as transposições dos espectros foram realizadas utilizando as estimativas para as notas mais graves obtidas na 1ª etapa do método 2C (avaliação completa). No segundo, as transposições dos espectros foram realizadas utilizando, sempre, a informação correta de qual é a nota mais grave de cada combinação (avaliação parcial). Na avaliação completa, o desempenho do método foi medido sobre todas as classificações obtidas (inclusive as das notas mais graves). Na avaliação parcial, o desempenho do método foi medido descontando as classificações das notas mais graves. Deste modo, foi possível avaliar o desempenho da 2ª etapa do método 2C, independentemente dos resultados da 1ª etapa.

Para formar os vetores-objetivo deste método, os elementos dos vetores-objetivo originais (vetores usados nos métodos 1A e 1B) são deslocados, de modo que o elemento referente à nota mais grave se torne, sempre, o primeiro elemento do vetor. Por exemplo, para um vetor-objetivo[3] original igual a $[0\ 0\ 1\ 0\ 0\ 1\ 0\ 1]^t$, o novo vetor-objetivo será $[1\ 0\ 0\ 1\ 0\ 1\ 0\ 0]^t$.

As redes desenvolvidas para a 2ª etapa do método 2C têm 234 neurônios na camada oculta, o mesmo número de elementos de seus vetores de entrada. Na Tabela 5.6 são mostrados os resultados da implementação desta versão do método 2C (avaliações parcial e completa), em 3 realizações, na classificação do conjunto de teste.

O melhor resultado foi obtido na 3ª realização, com uma rede treinada em

[3] Apesar de o exemplo apresentar vetores de 8 elementos, os vetores-objetivo, usados nos métodos de identificação de notas de violão, têm 44 elementos.

Tabela 5.6: Método 2C - 2ª etapa

rede	nº de épocas	acurácia (completa)	acurácia (parcial)
1	72	81,1	85,2
2	74	81,3	85,6
3	82	**81,5**	**85,7**

82 épocas. Outros resultados obtidos na avaliação completa são mostrados nas Figuras 5.9 e 5.10. Para esta realização, os resultados do NER e CER na avaliação completa foram, respectivamente, 15,1% e 42,3%. O resultado do NER obtido na avaliação parcial foi igual a 12,1%.

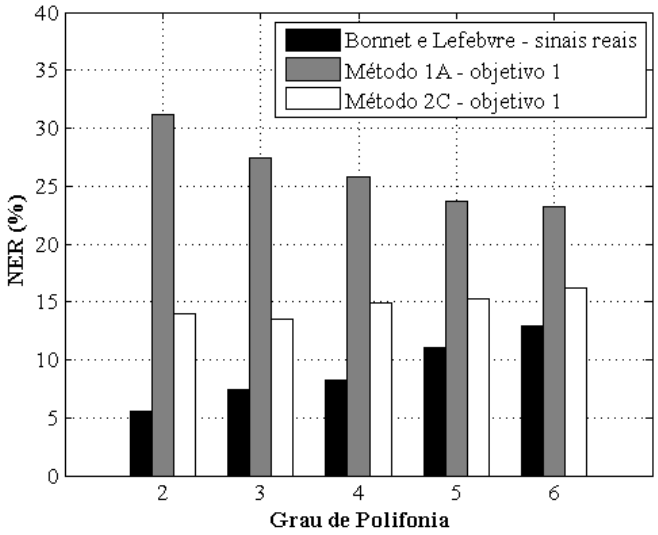

Figura 5.9: Percentuais do NER por grau de polifonia para o método de Bonnet e Lefebvre (na classificação de sinais reais de violão), para os métodos 1A e 2C (na classificação do conjunto de teste).

Apesar de 21,9% das estimativas da nota mais grave utilizadas nesta realização do método 2C estarem erradas, os valores do NER por grau de polifonia foram significativamente menores que os valores obtidos utilizando o método 1A. Isto ocorre porque, dado que a estimativa da nota mais grave esteja correta, a estimativa de intervalos realizada na 2ª etapa do método 2C tem melhor desempenho do que a estimativa direta de todas as notas, como no método 1A. Na Figura 5.11 estão os valores do NER por grau de polifonia para o método 1A e para o método 2C

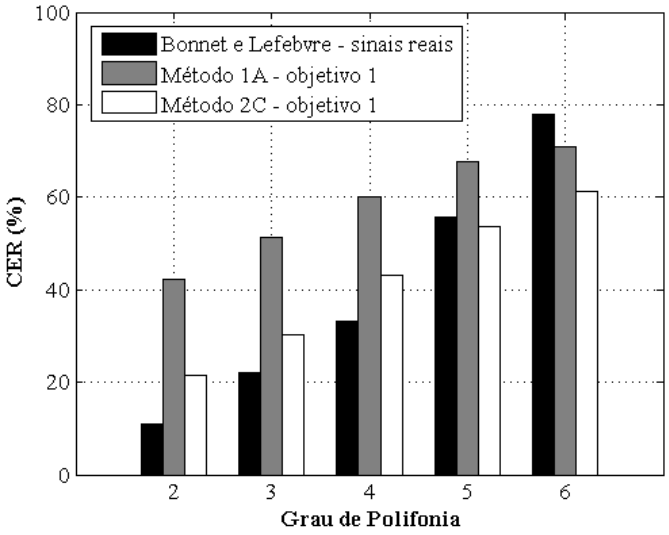

Figura 5.10: Percentuais do CER por grau de polifonia para o método de Bonnet e Lefebvre (na classificação de sinais reais de violão), para os métodos 1A e 2C (na classificação do conjunto de teste).

com avaliação parcial. A valor do NER por grau de polifonia diminuiu consideravelmente em comparação com os resultados do método 1A[4]. Além disto, muitas das estimativas erradas obtidas na 1ª etapa do método 2C, indicam notas que, apesar de não serem as mais graves, pertencem às combinações testadas.

5.2.3 Conclusões

Da comparação dos resultados obtidos a partir das aplicações dos métodos 1A e 1B, pode-se observar que o conhecimento do grau de polifonia de cada combinação analisada pode ser usado para reduzir o número de ocorrências de falsos positivos. Ao estabelecer que o número de notas estimadas em uma combinação deve ser igual ao seu grau de polifonia, impede-se a geração de erros de inserção.

A divisão do problema de identificação de notas musicais em duas etapas, sendo uma para a identificação da nota mais grave e outra para identificação dos intervalos entre a nota mais grave e as notas restantes, produziu melhores resultados do que a tentativa de estimar todas as notas simultaneamente. Isto ocorre porque, dado que a estimativa da nota mais grave esteja correta, a identificação dos intervalos

[4]Apesar de, na avaliação parcial, ser utilizada a informação correta de qual é a nota mais grave de cada combinação, estes acertos são descontados.

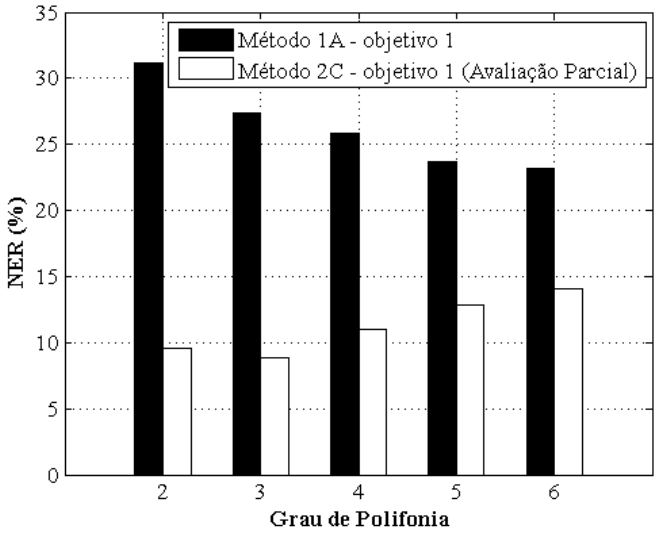

Figura 5.11: Percentuais do NER por grau de polifonia para o método 1A e para o método 2C (avaliação parcial), na classificação do conjunto de teste.

entre a nota mais grave e as notas restantes (realizada na 2ª etapa do método 2C) tem melhor desempenho do que a estimativa direta de todas as notas (realizada no método 1A). Além disto, mesmo quando erradas, as estimativas para as notas mais graves obtidas na 1ª etapa dos métodos do segundo grupo comumente indicam notas que também pertencem às combinações testadas. Nestes casos, mesmo com a decorrente falha na 2ª etapa, pelo menos uma nota é corretamente indicada. O melhor desempenho de identificação da nota mais grave foi obtido utilizando-se a 1ª etapa do método 2C.

5.3 Métodos para Identificação de Notas de Violão - Objetivo 2

Nesta seção são apresentadas adaptações de métodos descritos na Seção 5.2. Estas adaptações são voltadas para a identificação de notas em combinações nas quais uma nota tem dinâmica *forte* e as restantes têm dinâmica *mezzo*. Os segmentos dos registros usados para criar as combinações foram extraídos de trechos que compreendem, aproximadamente, o período de sustentação das notas. Foram realizadas adaptações dos métodos que não usam o conhecimento prévio do grau de polifonia.

As adaptações foram realizadas através da mudança dos conjuntos de treinamento, teste e validação, de acordo com as regras descritas na Seção 4.6 para o objetivo 2.

5.3.1 Método 1A - Objetivo 2

Os treinamentos das redes foram realizados de acordo com a metodologia apresentada na Seção 4.6. Foram treinadas 3 redes diferentes, todas com 234 neurônios na camada oculta, o mesmo número de elementos do vetor de entrada. Cada realização foi inicializada com um grupo de pesos sinápticos diferentes, cada um deles selecionado aleatoriamente de uma distribuição uniforme dos valores entre -0,25 e 0,25.

Os critérios de classificação utilizados nesta versão são os mesmos utilizados na Seção 5.2.1.1.

Na Tabela 5.7 são mostrados os resultados da implementação do método 1A, na classificação do conjunto de teste referente ao objetivo 2. O melhor resultado foi obtido na 2ª realização, com uma rede treinada em 75 épocas.

Tabela 5.7: Método 1A

rede	nº de épocas	acurácia %
1	81	66,9
2	75	**67,5**
3	63	66,2

Outros resultados desta classificação são mostrados nas Figuras 5.12 e 5.13, conjuntamente com os resultados da versão do método 1A desenvolvida para o objetivo 1.

Devido à presença de um nível extra de dinâmica ocorreram aumentos em quase todas as medidas de erro, exceto para o NER de 6 notas simultâneas. Pelo critério de classificação, o número de notas estimadas é no máximo igual a 6. Por isto não podem ocorrer erros de inserção para combinações de 6 notas (este tipo de erro só ocorre quando, para uma combinação, existirem mais falsos positivos que

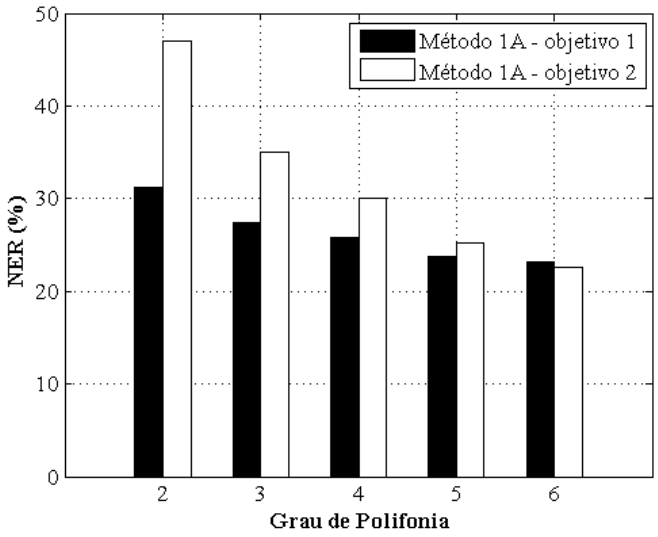

Figura 5.12: Percentuais do NER por grau de polifonia para versões do método 1A referentes aos objetivos 1 e 2.

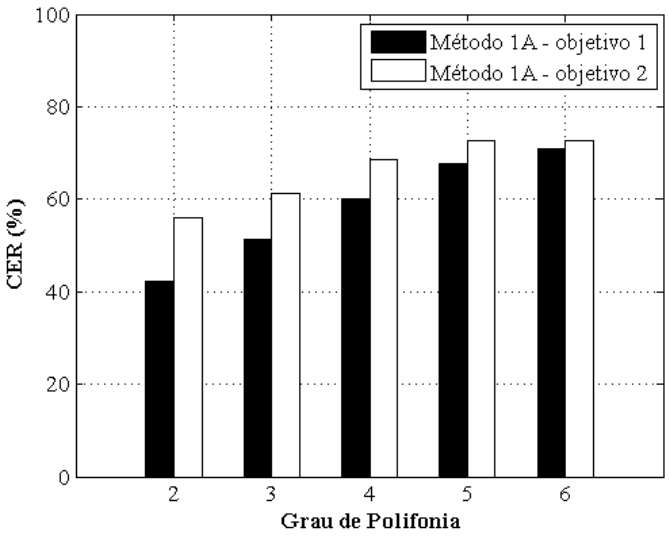

Figura 5.13: Percentuais do CER por grau de polifonia para versões do método 1A referentes aos objetivos 1 e 2.

notas-objetivo). Esta limitação para o número de notas estimadas contribui para a redução do NER para graus de polifonia mais altos.

Na Figura 5.14 são mostrados os percentuais de falsos positivos por grau de polifonia para as versões do método 1A referentes aos objetivos 1 e 2. Pode-se observar que o aumento no número de falsos positivos, que ocorre para o objetivo 2, é maior para combinações de poucas notas.

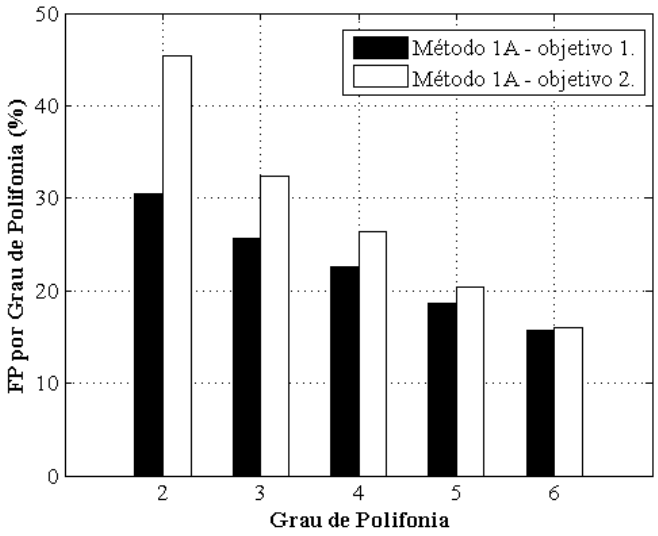

Figura 5.14: Percentuais de falsos positivos por grau de polifonia para versões do método 1A referentes aos objetivos 1 e 2.

Na Figura 5.15 são mostrados os percentuais de falsos negativos por dinâmica (a quantidade de falsos negativos para notas associadas a uma determinada dinâmica dividida pelo número de notas-objetivo associadas à mesma dinâmica). O percentual de notas que não foram encontradas é maior entre as notas com dinâmica *mezzo* do que entre as notas com dinâmica *forte*. Como só existe uma nota com dinâmica *forte* por combinação, suas parciais (normalmente com maior amplitude) se destacam nos espectros analisados.

5.3.2 Método 2A - 1ª etapa - Objetivo 2

Como os melhores resultados obtidos entre os métodos analisados na Seção 5.2.2, na etapa de identificação da nota mais grave, foram próximos (24,5% para o método 2A, 23,1% para o método 2B e 21,9% para o método 2C), esta etapa foi repetida para todos, com a troca dos conjuntos de treinamento, teste e validação,

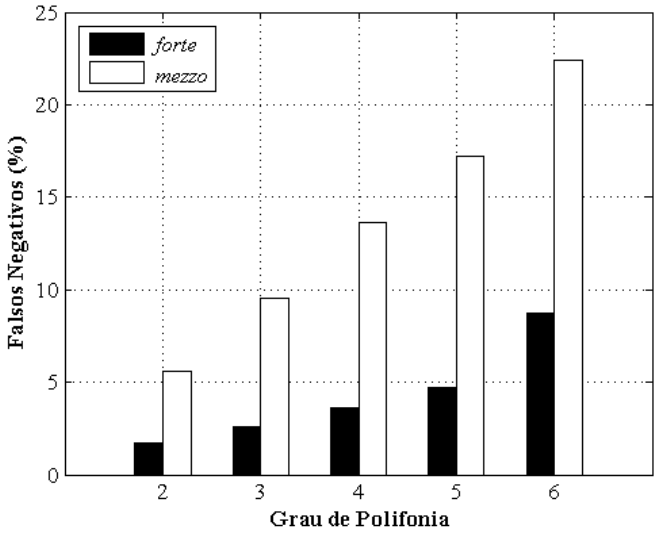

Figura 5.15: Percentuais de falsos negativos para dinâmicas *forte* e *mezzo*, para cada grau de polifonia, obtidos na aplicação do método 1A ao objetivo 2.

de acordo com o objetivo 2.

Os procedimentos de classificação utilizados nesta versão da 1ª etapa do método 2A são iguais aos apresentados na Seção 5.2.2.1.

Na Tabela 5.8 são mostrados os resultados da implementação desta versão para três realizações diferentes de redes com 234 neurônios na camada oculta.

Tabela 5.8: Método 2A - 1ª etapa

rede	nº de épocas	erro nmg
1	99	**24,2**
2	65	24,6
3	96	25,8

O melhor resultado obtido na aplicação da 1ª etapa do método 2A para o objetivo 2 (erro nmg=24,2%) foi muito próximo do melhor resultado obtido na aplicação do método desenvolvido para o objetivo 1 (erro nmg=24,5%).

5.3.3 Método 2B - 1ª etapa - Objetivo 2

Os critérios de classificação utilizados nesta versão da 1ª etapa do método 2B são iguais ao apresentados na Seção 5.2.2.2.

Na Tabela 5.9 são mostrados os resultados da implementação desta versão para três realizações diferentes, utilizando redes com 290 neurônios na camada oculta (a mesma quantidade de elementos dos vetores de entrada).

Tabela 5.9: Método 2B - 1ª etapa

rede	nº de épocas	erro nmg
1	81	22,2
2	75	**21,6**
3	63	24,1

O erro obtido na segunda realização do método 2B para o objetivo 2 (21,6%) foi menor que o erro mais baixo obtido na aplicação deste método desenvolvida para o objetivo 1 (23,1%).

5.3.4 Método 2C - 1ª etapa - Objetivo 2

Os procedimentos de classificação utilizados nesta versão são iguais aos apresentados na Seção 5.2.2.3.

Na Tabela 5.10 são mostrados os resultados da implementação desta versão para três realizações diferentes, utilizando redes com 290 neurônios na camada oculta (a mesma quantidade de elementos dos vetores de entrada).

Foram treinadas 3 redes diferentes para realizar a classificação, todas com 290 neurônios na camada oculta. Na Tabela 5.10 são mostrados os resultados da implementação da 1ª etapa do método 2C.

Tabela 5.10: Método 2C - 1ª etapa

rede	nº de épocas	erro nmg
1	60	**19,4**
2	58	21,3
3	53	20,4

Novamente, este método obteve o melhor desempenho entre os métodos da 1ª etapa (erro nmg = 19,4%).

O erro obtido na segunda realização do método 2C para o objetivo 2 foi melhor que o erro mais baixo obtido na aplicação deste método desenvolvida para o objetivo 1 (21,9%). Nas duas versões apresentadas para a 1ª etapa dos métodos 2A, 2B e 2C o menor erro foi obtido com o método 2C.

5.3.5 Método 2C - 2ª etapa - Objetivo 2

As redes desenvolvidas para a 2ª etapa do método 2C têm 234 neurônios na camada oculta, o mesmo número de elementos de seus vetores de entrada. Na Tabela 5.11 são mostrados os resultados da implementação desta versão do método 2C (avaliações parcial e completa), em 3 realizações, na classificação do conjunto de teste.

Os critérios de classificação utilizados nesta etapa do método 2C são iguais aos apresentados na Seção 5.2.2.4.

Tabela 5.11: Método 2C - 2ª etapa

rede	nº de épocas	acurácia (completa)	acurácia (parcial)
1	105	78,1	79,0
2	105	**78,5**	**79,3**
3	99	77,8	78,5

O melhor resultado foi obtido na 2ª realização, com uma rede treinada em 105 épocas. Para esta realização, os resultados do NER e CER na avaliação completa foram, respectivamente, 18,2% e 53,1%. O resultado do NER obtido na avaliação parcial foi igual a 18,2%. Do total de falsos negativos, apenas 1,2% ocorreram para notas com dinâmica *forte*.

As medidas de NER e CER em cada grau de polifonia na avaliação completa são mostradas na Figuras 5.16 e 5.17, em conjunto com as medidas obtidas por Bonnet e Lefebvre (apresentadas anteriormente na Seção 5.2). Deve-se destacar que Bonnet e Lefebvre não apresentaram informações sobre a dinâmica das notas presentes em seu banco de sinais reais de violão. Seus resultados são mostrados nesta seção apenas como uma referência.

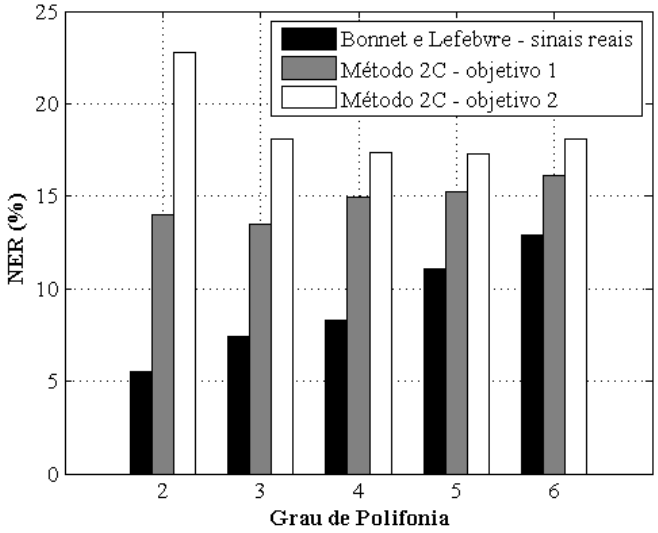

Figura 5.16: Percentuais do NER por grau de polifonia para o método de Bonnet e Lefebvre (na classificação de sinais reais de violão) e para duas versões do método 2C (nas classificações referentes aos objetivos 1 e 2).

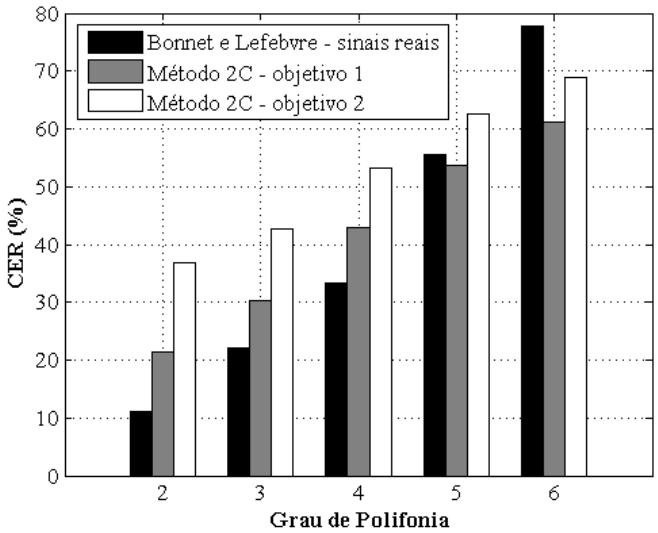

Figura 5.17: Percentuais do CER por grau de polifonia para o método de Bonnet e Lefebvre (na classificação de sinais reais de violão) e para duas versões do método 2C (nas classificações referentes aos objetivos 1 e 2).

Apesar de o percentual de estimativas erradas para a nota mais grave utilizado nesta implementação do método 2C (19,4%), ser menor que o percentual obtido na implementação referente ao objetivo 1 (21,9%), a presença de apenas uma nota com dinâmica *forte* por combinação provocou o aumento dos percentuais de NER e CER para todos os graus de polifonia.

5.3.6 Conclusões

Para a versão do método 1A referente ao objetivo 2, a presença de uma nota com dinâmica *forte* por combinação provocou (em comparação com os resultados obtidos para o objetivo 1) o aumento do NER para todos os graus de polifonia analisados, exceto para 6 notas simultâneas. O limite para o número de notas estimadas, estabelecido no critério de classificação, impede a ocorrência de erros de inserção na análise de combinações de 6 notas e contribui para a redução do NER nas classificações de combinações com número de notas próximo ao grau de polifonia do instrumento.

O percentual de notas que não foram encontradas (falsos negativos) foi, para os métodos 1A e 2C, maior entre as notas com dinâmica *mezzo* do que entre as notas com dinâmica *forte*.

Novamente, o melhor desempenho de identificação da nota mais grave foi obtido utilizando-se a 1ª etapa do método 2C.

Assim como na comparação das versões dos métodos 1A e 2C desenvolvidas para o objetivo 1, no desenvolvimento para o objetivo 2 a divisão do problema em uma etapa de identificação da nota mais grave e outra para identificação dos intervalos entre a nota mais grave e as notas restantes teve melhores resultados do que a tentativa de estimar todas as notas simultaneamente.

5.4 Métodos para Identificação de Notas de Violão - Objetivo 3.

Nesta seção são apresentadas adaptações dos métodos 1A e 2C. Estas adaptações são voltadas para a identificação de notas em combinações de registros com dinâmica *mezzo* a partir de três possibilidades de segmentação: todos os segmentos

extraídos aproximadamente do período que compreende o ataque e decaimento, todos os segmentos extraídos aproximadamente do período de sustentação e todos os segmentos extraídos aproximadamente do período de liberação.

5.4.1 Método 1A - Objetivo 3

Os treinamentos das redes utilizadas nesta adaptação foram realizados de acordo com a metodologia apresentada na Seção 4.6 para o objetivo 1. Os critérios de classificação utilizados são os mesmos apresentados na Seção 5.2.1.1.

Na Tabela 5.12 são mostrados os resultados da implementação do método 1A, na classificação do conjunto de teste referente ao objetivo 3. Foram treinadas 3 redes diferentes, todas com 234 neurônios na camada oculta. As medidas de acurácia obtidas para os diferentes segmentos são indicadas pelas letras[5] AD (ataque e decaimento), S (sustentação) e R (liberação). Os resultados foram muito próximos nas 3 realizações (os melhores estão destacados em negrito). Os piores resultados ocorreram, em todas as realizações, na classificação das combinações formadas com segmentos que compreendem os períodos de ataque e decaimento.

Tabela 5.12: Método 1A

rede	nº de épocas	acurácia % (AD)	acurácia % (S)	acurácia % (R)
1	38	63,6	70,2	70,1
2	36	63,5	**70,8**	**70,4**
3	43	**63,7**	70,1	69,3

A Figura 5.18 contém, para os períodos AD e S, o número de falsos positivos que ocorreram para cada nota. A quantidade de falsos positivos, para ambas as análises, é maior nas regiões correspondentes aos *pitches* que recaem sobre as faixas da 1ª e 2ª freqüências de ressonância do violão. Pode-se observar que a quantidade de falsos positivos na região da 1ª freqüência de ressonância é maior para o período que compreende aproximadamente o ataque e o decaimento. Este tipo de erro aumenta

[5]As letras utilizadas compõem a sigla, comumente usada, da modelagem de envoltória *Attack, Decay, Sustain and Release*.

por causa da presença mais acentuada da 1ª freqüência de ressonância durante este período, como descrito na Seção 1.7.

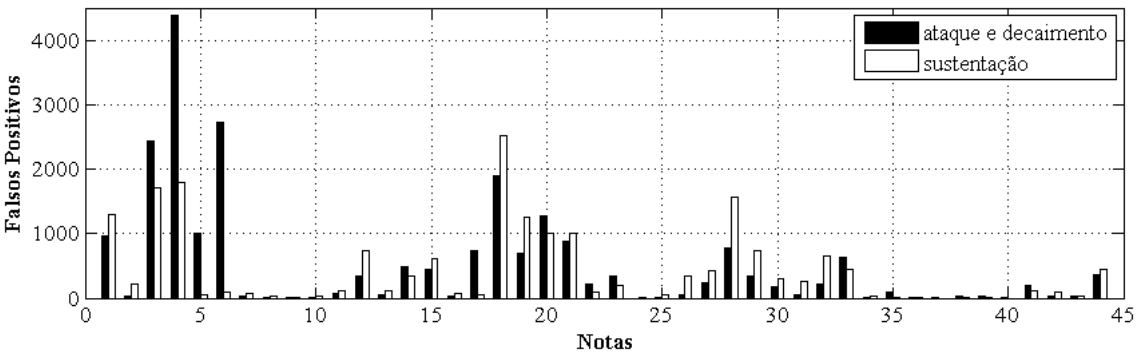

Figura 5.18: Histograma de Falsos Positivos.

5.4.2 Método 2C - 1ª etapa - Objetivo 3

Na Tabela 5.13 são mostrados os resultados da adaptação da 1ª etapa do método 2C ao objetivo 3.

Tabela 5.13: Método 2C - 1ª etapa

rede	nº de épocas	erro nmg % (AD)	erro nmg % (S)	erro nmg % (R)
1	35	39,4	21,5	15,8
2	34	35,8	20,7	16,8
3	28	**35,5**	**19,5**	**15,3**

Nas 3 realizações, o número de erros na estimação da nota mais grave é maior no período que compreende aproximadamente o ataque e o decaimento, e decai para os períodos seguintes. Este efeito é mais destacado nesta etapa do que nas estimativas realizadas na Seção 5.4.1. Isto ocorre porque muitas das notas mais graves têm parciais (dadas pela Equação (1.2)) com frequências próximas a pelo menos uma das freqüências de ressonância mais baixas do instrumento. No período de ataque, o acoplamento entre estas parciais e as freqüências de ressonância é mais acentuado do que nos períodos seguintes.

5.4.3 Método 2C - 2ª etapa - Objetivo 3

Na Tabela 5.14 são mostrados os resultados da implementação desta versão do método 2C (avaliações parcial e completa), em 3 realizações, na classificação do conjunto de teste.

Tabela 5.14: Método 2C - 2ª etapa

	nº de	acurácia (AD)		acurácia (S)		acurácia (R)	
rede	épocas	total	parcial	total	parcial	total	parcial
1	84	74,7	81,7	**83,1**	86,1	80,9	81,4
2	77	74,7	81,5	83,0	86,5	80,5	81,0
3	100	**74,9**	**82,1**	**83,1**	**86,2**	**81,0**	**81,6**

O melhor resultado foi obtido na 3ª realização, com uma rede treinada em 100 épocas. Para esta realização, os resultados do NER e CER (avaliação completa) dos vetores referentes ao período que compreende aproximadamente o ataque e decaimento, foram, respectivamente, 21,6% e 60,1%. O resultado do NER obtido na avaliação parcial foi igual a 15,7%. Os resultados do NER e CER na avaliação completa para o período de sustentação, foram, respectivamente, 13,7% e 38,8%. O resultado do NER obtido na avaliação parcial foi igual a 12,1%. Os resultados do NER e CER na avaliação completa para o período de liberação, foram, respectivamente, 15,4% e 45,0%. O resultado do NER obtido na avaliação parcial foi igual a 15,5%.

Outros resultados obtidos na avaliação completa são mostrados nas Figuras 5.19 e 5.20.

A classificação de trechos que compreendem o ataque das notas é dificultada pela grande quantidade de modos presentes neste período (devidos à natureza impulsiva do plectro), como discutido na Seção 1.7. A classificação trechos extraídos aproximadamente do período de liberação pode ser dificultada pela redução da razão sinal/ruído. Os melhores resultados, em todos as medidas, foram obtidos na classificação de trechos extraídos aproximadamente do período de sustentação.

Todos vetores referentes ao objetivo 1 e parte dos vetores referentes ao objetivo 3 compartilham as mesmas características: foram criados usando notas com

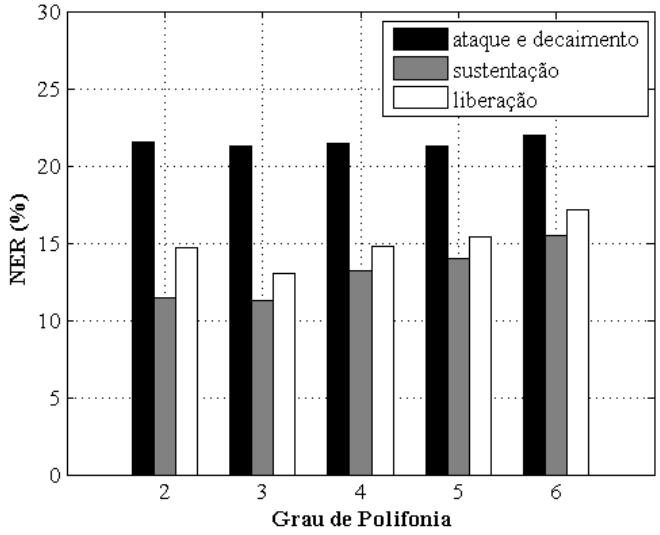

Figura 5.19: Percentuais do NER por grau de polifonia para o método 2C para classificações referentes as segmentações sobre os períodos, aproximados, de ataque e decaimento, sustentação e liberação.

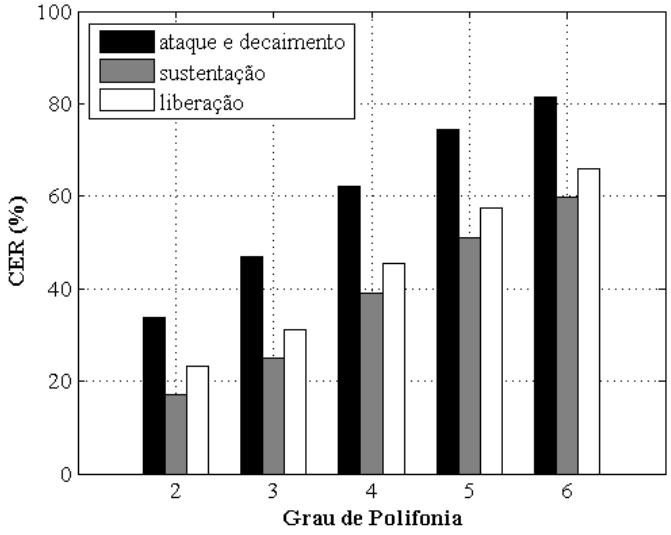

Figura 5.20: Percentuais do CER por grau de polifonia para o método 2C para classificações referentes as segmentações sobre os períodos, aproximados, de ataque e decaimento, sustentação e liberação.

dinâmica *mezzo* e trechos obtidos a partir do período de sustentação. Os melhores resultados para os métodos 1A e 2C referentes ao objetivo 1 foram, respectivamente, 69,7% e 81,5%; já os melhores resultados para os métodos 1A e 2C referentes ao objetivo 3 para vetores com as mesmas características foram, respectivamente, 70,8% e 83,1%. Isto indica que aumentar a variabilidade de exemplos de treinamento aumenta a robustez da análise.

5.4.4 Conclusões

Para esta versão do método 1A, as menores acurácias foram obtidas nas classificações de vetores criados a partir de segmentos extraídos, aproximadamente, dos períodos de ataque e decaimento. Para estes períodos, e para o período de sustentação, a maior parte dos casos de falsos positivos foram obtidos nas regiões correspondentes aos *pitches* que recaem sobre as faixas da 1ª e 2ª freqüências de ressonância do violão. Particularmente, pode-se observar que a quantidade de falsos positivos na região da 1ª freqüência de ressonância é maior no período que compreende aproximadamente o ataque e o decaimento, resultado da presença acentuada da 1ª freqüência de ressonância durante este período.

Novamente o método 2C gerou melhores resultados que o método 1A (neste caso para todas as segmentações).

Os resultados dos métodos 1A e 2C, obtidos da análise de vetores criados a partir de segmentos com dinâmica *mezzo* extraídos do período de sustentação das notas, foram melhores que os resultados dos métodos desenvolvidos para o objetivo 1, também obtidos da análise de vetores criados a partir de segmentos com dinâmica *mezzo* extraídos do período de sustentação das notas. Isto serve como indicação de que o aumento da variedade de exemplos de treinamento (utilizados nos métodos referentes ao objetivo 3) aumenta a robustez da análise.

Os melhores resultados do método 2C foram obtidos na classificação de trechos extraídos aproximadamente do período de sustentação. A classificação de trechos que compreendem o ataque das notas é dificultada pela grande quantidade de modos presentes neste período (devidos à natureza impulsiva do plectro). A classificação de trechos extraídos aproximadamente do período de liberação pode ser dificultada pela redução da razão sinal/ruído.

Deve-se ressaltar que as combinações de notas utilizadas para desenvolver esta dissertação foram realizadas computacionalmente. Desta forma, as amplitudes de parciais resultantes de acoplamentos entre modos de diferentes cordas e dos tampos inferior e superior do violão podem não ter sido bem aproximadas. A combinação automática pode, por exemplo, gerar parciais sobre as freqüências de ressonância com amplitude consideravelmente maior do que seria encontrada em um registro do instrumento que contenha as mesmas notas. Bancos de dados com registros de acordes realizados por um músico devem ser testados em trabalhos futuros.

5.5 Métodos para Identificação de Notas de Violão - Objetivo 4.

Nesta seção são apresentadas adaptações dos métodos 1A e 2C voltadas para a identificação de notas em combinações de registros com dinâmicas escolhidas aleatoriamente entre *forte*, *mezzo* e *piano*, extraídos de períodos escolhidos aleatoriamente entre ataque e decaimento, sustentação e liberação.

O conjunto dos vetores criados para os experimentos do objetivo 4 simulam situações mais complexas do que as abordadas nos experimentos referentes aos primeiros 3 objetivos. As notas presentes em um instante qualquer de uma gravação real podem estar em etapas diferentes na evolução de suas envoltórias e, simultaneamente, terem diferentes dinâmicas.

5.5.1 Método 1A - Objetivo 4

Os treinamentos das redes utilizadas nesta adaptação foram realizados de acordo com a metodologia apresentada na Seção 4.6 para o objetivo 1. Os critérios de classificação utilizados são os mesmos apresentados na Seção 5.2.1.1.

Na Tabela 5.15 são mostrados os resultados da implementação do método 1A com 3 redes diferentes de 234 neurônios na camada oculta, utilizando 44504 vetores no conjunto de treinamento e no conjunto de validação.

Para aprimorar os resultados, além dos experimentos com resultados apresentados na Tabela 5.15, foram testadas realizações com diferentes topologias de rede e quantidades de vetores para treinamento e validação. As mudanças de topologia

Tabela 5.15: Método 1A - Objetivo 4

rede	nº de épocas	acurácia
1	72	58,4
2	67	**58,9**
3	62	58,1

foram obtidas alterando o número de neurônios na camada oculta. Foram realizados testes utilizando topologias com 209, 184, 159 e 134 neurônios na camada oculta. Para cada topologia foram feitos testes com 20504, 26504, 32504, 38504, 44504, 50504 e 56504 pares de vetores de treinamento. O melhor resultado (acurácia igual a **59,3%**), foi obtido utilizando uma rede com 209 neurônios na camada oculta e 44504 pares de vetores para os conjuntos de treinamento e validação. Os valores de NER e CER foram, respectivamente, 40,8% e 83,2 %, percentuais bem mais elevados do que os encontrados na implementação referente ao objetivo 1 (NER=25,3% e CER=58,7%), onde todas as notas nas combinações tinham a mesma dinâmica e foram extraídas do período de sustentação.

Do total de falsos negativos desta implementação, 71,9% ocorreram para notas com dinâmica *piano*, 23,4% para notas com dinâmica *mezzo* e apenas 5,1% para notas com dinâmica *forte*. O total de falsos negativos em função dos períodos aproximados de segmentação foi dividido em 34,1% para notas segmentadas a partir do ataque e decaimento, 31,7% para notas segmentadas a partir do período de sustentação e 34,2% para notas segmentadas a partir do período de liberação.

Outros resultados desta implementação são mostrados nas Figuras 5.21 e 5.22, conjuntamente com os resultados da versão do método 1A desenvolvida para o objetivo 1.

5.5.2 Método 2C - 1ª etapa - Objetivo 4

Na Tabela 5.16 são mostrados os resultados da implementação da ª etapa do método 2C para o objetivo 4, em 3 realizações diferentes. Foram utilizadas redes com 234 neurônios na camada oculta, utilizando 44504 vetores no conjunto de treinamento e no conjunto de validação.

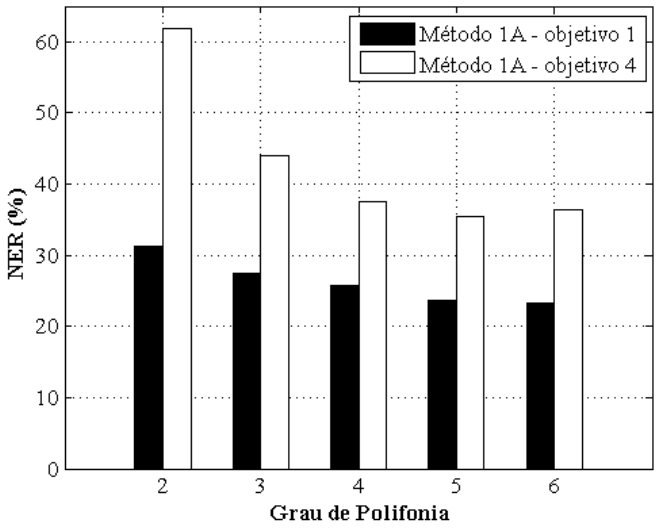

Figura 5.21: Percentuais do NER por grau de polifonia para versões do método 1A referentes aos objetivos 1 e 4.

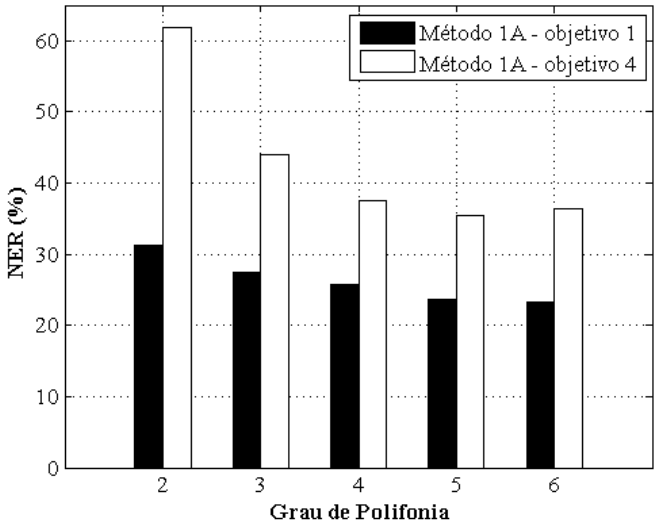

Figura 5.22: Percentuais do CER por grau de polifonia para versões do método 1A referentes aos objetivos 1 e 4.

Tabela 5.16: Método 2C - 1ª etapa

rede	nº de épocas	erro nmg
1	72	37,5
2	67	**36,5**
3	62	37,5

Novamente, para aprimorar os resultados, além dos experimentos com resultados apresentados na Tabela 5.16, foram testadas realizações com diferentes topologias de rede e quantidades de vetores para treinamento e validação. As alternativas testadas para topologia e quantidade de vetores utilizados no treinamento foram as mesmas descritas na Seção 5.5.1. O melhor resultado (**35,9%**), foi obtido utilizando uma rede com 234 neurônios na camada oculta e 38504 pares de vetores para os conjuntos de treinamento e validação.

A maior parte dos erros, 62,9% do total, ocorreu na identificação de notas com dinâmica *piano*. O resto dos erros foi dividido em 20,5% para notas com dinâmica *mezzo* e 16,6% para notas com dinâmica *forte*. O total de falsos negativos em função dos períodos aproximados de segmentação foi dividido em 37,3% para notas segmentadas a partir do ataque e decaimento, 30,9% para notas segmentadas a partir do período de sustentação e 31,8% para notas segmentadas a partir do período de liberação.

5.5.3 Método 2C - 2ª etapa - Objetivo 4

Na Tabela 5.17 são mostrados os resultados da implementação do método 2C com 3 redes diferentes de 234 neurônios na camada oculta, utilizando 44504 vetores no conjunto de treinamento e no conjunto de validação. Foram utilizadas as estimativas com o menor percentual de erros, 35,92%, obtidas na implementação da 1ª etapa do método 2C.

Tabela 5.17: Método 1A - Objetivo 4

rede	nº de épocas	acurácia (completa)	acurácia (parcial)
1	138	66,1	68,6
2	134	66,1	68,3
3	154	**66,2**	**68,7**

Seguindo o mesmo procedimento apresentado para o método 1A e para a 1ª etapa do método 2C, foram testadas realizações com diferentes topologias de rede e quantidades de vetores para treinamento e validação. As alternativas foram as mesmas descritas na Seção 5.5.1.

O melhor resultado, acurácia igual a **66,2%**, foi obtido utilizando uma rede com 234 neurônios na camada oculta e 54504 pares de vetores para os conjuntos de treinamento e validação. Os valores de NER e CER obtidos na análise completa foram, respectivamente, 30,4% e 74,6%. Na análise parcial os valores obtidos para a acurácia e para o NER foram, respectivamente, 68,8% e 21,0%. Outros resultados desta classificação são mostrados nas Figuras 5.23 e 5.24, conjuntamente com os resultados da versão do método 2C desenvolvida para o objetivo 1.

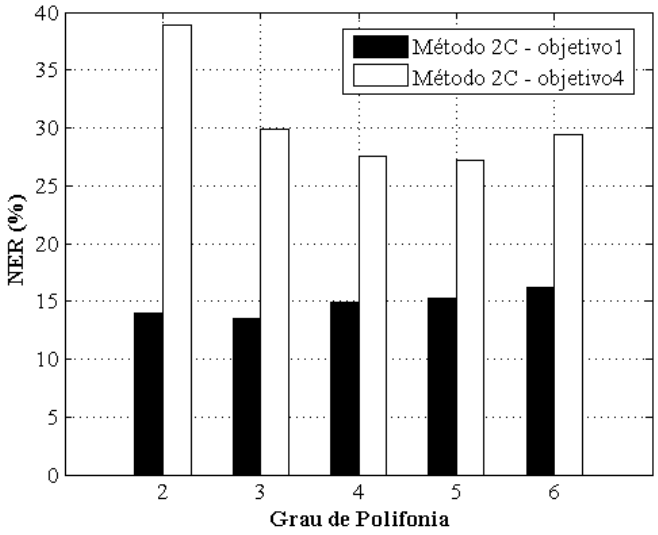

Figura 5.23: Percentuais do NER por grau de polifonia para as versões do método 2C referentes aos objetivos 1 e 4.

Do total de falsos negativos desta implementação, 72,8% ocorreram para notas com dinâmica *piano*, 21,7% para notas com dinâmica *mezzo* e 5,5% para notas com dinâmica *forte*. O total de falsos negativos em função dos períodos aproximados de segmentação foi dividido em 33,7% para notas segmentadas a partir do ataque e decaimento, 31,1% para notas segmentadas a partir do período de sustentação e 35,2% para notas segmentadas a partir do período de liberação.

5.5.4 Conclusões

Tanto para o método 1A quanto para o método 2C, a maior parte das notas que não foram encontradas tinham dinâmica *piano*. O segundo maior percentual de falsos negativos ocorreu na classificação de notas com dinâmica *mezzo* e o menor, na classificação de notas com dinâmica *forte*.

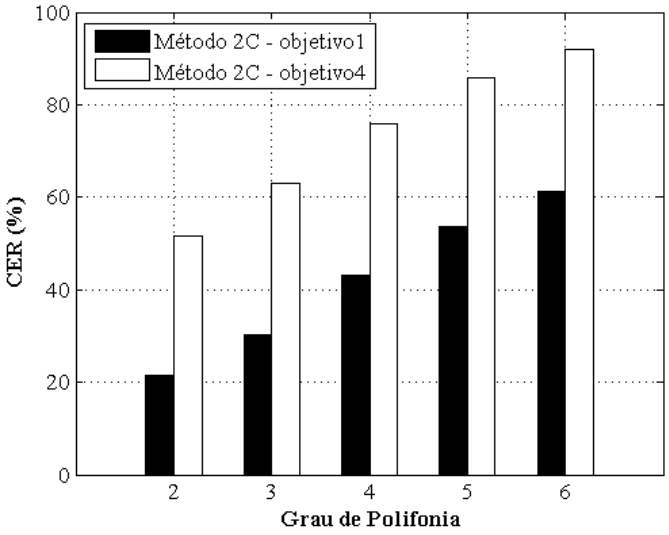

Figura 5.24: Percentuais do CER por grau de polifonia para as versões do método 2C referentes aos objetivos 1 e 4.

Não ocorreram diferenças elevadas nos percentuais de falsos positivos para cada período de segmentação. Para ambos os métodos, cada período recebeu acima de 31% do total de falsos negativos. Os menores percentuais, porém, ocorreram na identificação de notas extraídas aproximadamente do período de sustentação.

Os resultados obtidos com os métodos 1A e 2C voltados para o objetivo 4 foram inferiores aos resultados obtidos com versões destes métodos voltadas para os 3 primeiros objetivos. Isto ocorreu porque o conjunto de vetores criados para os experimentos do objetivo 4 simulam situações mais complexas (com dinâmicas e segmentações escolhidas aleatoriamente) do que as abordadas nos experimentos referentes aos outros objetivos.

O método 2C gerou, assim como nas análises referentes aos objetivos 1, 2 e 3, melhores resultados que o método 1A.

Capítulo 6

Metodologia para Identificação de Notas de Piano

6.1 Introdução

Neste capítulo e no capítulo seguinte são apresentadas adaptações dos métodos dos Capítulos 4 e 5, voltadas para a identificação de notas musicais em registros de piano solo.

Os métodos para identificação de notas de piano seguem os mesmos princípios dos métodos apresentados para identificação de notas de violão. São divididos em dois grupos principais: no primeiro, métodos que utilizam apenas uma rede neural e apenas uma representação espectral na identificação das notas de cada combinação; no segundo, métodos que utilizam duas redes neurais e duas representações espectrais na identificação das notas de cada combinação.

Os métodos do segundo grupo para identificação de notas de piano, assim como os métodos do segundo grupo para identificação de notas de violão, têm uma rede para identificar a nota mais grave de cada combinação e outra para encontrar os intervalos entre a nota mais grave e as notas restantes. Da mesma forma que foi apresentada no Capítulo 4, a segunda rede recebe como vetor de entrada uma versão 'transposta' da representação espectral. As duas redes são utilizadas em seqüência. Após conhecer a estimativa para a nota mais grave, o espectro da CQT é alterado para que a componente analisada sobre o *pitch* da nota mais grave se torne a primeira componente do espectro. O novo espectro é então analisado com

a segunda rede para obter estimativas dos intervalos entre a nota mais grave da combinação e as notas restantes.

6.2 Banco de Dados

Os registros de notas de piano foram obtidos do banco de gravações de áudio *RWC Music Database: Musical Instrument Sound Database* [37] e do banco de gravações de áudio *McGill University Master Samples* [44]. O banco de gravações de piano RWC é composto por registros digitais monaurais com resolução de 16 bits e taxa de amostragem de 44100 Hz. O banco de gravações de piano MUMS[1] é composto por registros digitais estéreo com resolução de 32 bits e taxa de amostragem de 44100 Hz.

As gravações de piano do banco RWC utilizadas nesta tese, nomeadas com as siglas 011PFNOP, 011PFNOM, 011PFNOF, 012PFNOP, 012PFNOM, 012PFNOF, 013PFNOP, 013PFNOM, 013PFNOF, 011PFSTP, 011PFSTM, 011PFSTF, 012PFSTP, 012PFSTM, 012PFSTF, 013PFSTP, 013PFSTM e 013PFSTF, contêm seqüências de notas de 3 pianos diferentes (designados por '011PF', '012PF' e '013PF'). As gravações com nomes que contêm as letras 'NO' foram realizadas com técnica normal. As gravações com nomes que contêm as letras 'ST' foram realizadas com técnica *staccato*. Foram utilizados 3 níveis diferentes de dinâmica (indicados pela última letra de cada sigla, 'P' para *piano*, 'M' para *mezzo* e 'F' para *forte*). Cada gravação é composta por uma seqüência de 88 sons de notas individuais, cada som de uma nota diferente.

Os registros de cada nota de piano do banco MUMS são disponibilizados como gravações independentes, já segmentadas. As gravações são organizadas em grupos de 88 registros de notas diferentes. Os registros utilizados nesta tese pertencem aos grupos MPP SOFT, MPP MEDIUM, MPP LOUD, todos com sons de um mesmo piano. Cada grupo contém gravações com um nível diferente de dinâmica (indicados por 'SOFT' para *piano*, 'MEDIUM' para *mezzo* e 'LOUD' para *forte*). Foi utilizado apenas o canal esquerdo de cada gravação.

[1] A partir deste ponto, por brevidade, o banco *McGill University Master Samples* será referido apenas como banco MUMS.

6.3 Segmentação

A segmentação dos registros de piano do banco de dados RWC foi realizada através da análise visual das formas de onda de cada gravação. Cada registro foi disponibilizado no banco de dados como uma gravação independente, já segmentada.

Os registros de piano do banco de dados MUMS não requereram segmentação.

As marcações de *onsets* da base RWC utilizadas nesta teste estão listadas no Apêndice A.2.

6.4 Criação dos *Kernels* da CQT

A abordagem utilizada para a criação dos kernels descrita na Seção 4.4 foi repetida, exceto pela extensão da faixa de análise. Para o primeiro e segundo grupo de métodos foram criados *kernels* da CQT para a análise de componentes a partir do *pitch* da nota mais grave do piano (Lá 0; *pitch* = 27,50 Hz) até aproximadamente 21096,16 Hz. Com esta escolha para o limite superior é possível representar cinco parciais da nota mais aguda do instrumento (Dó 8; *pitch* ≈ 4186,01 Hz).

Novamente os *kernels* da 1^a oitava da transformada foram criados com resolução freqüencial $q = 2^{\frac{1}{12}}$, correspondente a 1 semitom. Deste modo, a duração do intervalo necessário para o cálculo da componente sobre o *pitch* de Lá 0 é aproximadamente igual a 0,61 s. Foram criados dois grupos de *kernels* para a análise da 2^a oitava: o 1^o com resolução freqüencial $q = 2^{\frac{1}{24}}$, correspondente a 1/4 de tom e o 2^o com resolução freqüencial $q = 2^{\frac{1}{36}}$, correspondente a 1/6 de tom. O 1^o grupo abrange, com 14 componentes, as 7 primeiras notas dessa oitava. O 2^o grupo abrange, com 15 componentes, as 5 últimas notas dessa oitava. A partir da 3^a oitava os *kernels* da transformada foram realizados com resolução freqüencial $q = 2^{\frac{1}{48}}$, correspondente a 1/8 de tom. No total, uma representação freqüencial criada utilizando estes *kernels* contém 406 componentes: 12 na 1^a oitava, 29 na 2^a oitava e 365 a partir da 1^a componente da 3^a oitava até a última componente da transformada.

Os *kernels* complementares para o cálculo dos espectros transpostos são calculados da mesma forma, acompanhando o aumento da faixa de análise. As transformadas com a primeira componente a partir do *pitch* da nota Lá♯ 0 (*pitch* ≈ 29,14 Hz), a segunda nota do instrumento, já teriam componentes calculadas acima da freqüên-

cia de Nyquist [39]. Os *kernels* referentes a estas componentes não são calculados e os valores de suas componentes são preenchidos com zeros durante o cálculo das transformadas.

6.5 Criação das Combinações de Notas Musicais

As combinações de notas foram novamente realizadas computacionalmente, criando sons com diferentes graus de polifonia, com até dez notas simultâneas. Assim como nas combinações de registros de violões, os sons gerados com este procedimento não apresentam os efeitos de acoplamento entre modos de vibração de cordas diferentes, que podem ocorrer durante a execução do instrumento.

A construção dos conjuntos de combinações obedeceu as normas descritas a seguir:

a. Para cada instrumento do banco de dados foi gerada uma série independente de combinações dos registros disponíveis. A escolha dos registros utilizados foi feita da seguinte forma:

1. As combinações com duas notas simultâneas foram escolhidas utilizando todas as combinações possíveis de dois registros de notas, exceto as combinações de registros com notas iguais. Assim foram criadas 3828 combinações por piano de ambas as bases de dados.

2. Na escolha dos registros de cada combinação com grau de polifonia maior que dois, o primeiro registro era escolhido aleatoriamente, de uma distribuição uniforme, entre os 88 registros disponíveis. Este registro então era retirado das opções disponíveis para a próxima escolha. O segundo registro era escolhido aleatoriamente, de uma nova distribuição uniforme, entre os registros restantes. Novamente o registro escolhido era retirado das opções disponíveis para a próxima escolha. Este processo era repetido até se completar o grau de polifonia desejado. Foram escolhidas 1000 combinações diferentes por cada grau de polifonia (a partir de três notas), por piano utilizado nos grupos de treino e validação. A redução na quantidade de combinações testadas, em comparação com as quantidades utilizadas nos testes de violão, se deu por limitações computacionais. Com o aumento no tamanho das representações espectrais e

no número de possíveis notas simultâneas, o consumo de memória durante o treinamento cresceu, dificultando o uso de mais combinações por instrumento. Para o grupo de teste foram escolhidos 2750 vetores por cada grau de polifonia (a partir de três notas), por piano utilizado. Como este grupo não é utilizado durante o treinamento, não tem influência no aumento do uso de memória durante a adaptação da rede. Por isto foi possível utilizar mais vetores.

b. Trechos diferentes dos registros escolhidos foram selecionados aleatoriamente, entre 3 opções: trechos que deveriam compreender aproximadamente o período de ataque e decaimento (segmentados a partir da primeira amostra do registro), trechos que deveriam compreender aproximadamente o período de sustentação (segmentados a partir da amostra 8001) e trechos que deveriam compreender aproximadamente o período de liberação (segmentados a partir da amostra 16001). As notas agudas do piano comumente têm períodos muitos curtos de sustentação e liberação, por isto, para associar valores coerentes com os inícios destes períodos, foram utilizados valores menores que os utilizados na segmentação dos registros de violão. Deste modo, reduz-se também o risco de realizar cálculos utilizando amostras localizadas após o período de liberação das notas agudas. Uma exceção foi usada na segmentação dos registros 012PFNO. A amostra 12001 foi associada ao início do período de liberação para esses registros. A redução foi devida a uma falha em alguns dos registros da gravação 012PFNOM: algumas notas têm a gravação emudecida precocemente. Todos os trechos foram segmentados com a duração do maior intervalo necessário para o cálculo da CQT, aproximadamente 0,61 s.

c. Antes de compor cada combinação, as dinâmicas dos segmentos utilizados eram escolhidas aleatoriamente, entre *forte*, *mezzo* e *piano*. Os trechos utilizados eram normalizados pela norma quadrática e em seguida, de acordo com a dinâmica escolhida, poderiam ter suas amplitudes alteradas. Quando a dinâmica escolhida era *forte*, a amplitude era mantida. Quando a dinâmica escolhida era *mezzo*, a amplitude era alterada, formando sinais com $-10\,\mathrm{dB}$ de potência em relação aos segmentos normalizados. Quando a dinâmica escolhida era *piano*, a amplitude era alterada, formando sinais com $-20\,\mathrm{dB}$ de potência em relação aos segmentos normalizados. Ambas as bases possuem gravações com níveis de dinâmica *forte*, *mezzo* e *piano*. Os registros eram selecionados entre as opções de dinâmica disponíveis (listadas na

Seção 6.2) de acordo com as dinâmicas escolhidas.

d. Sinais de notas simples também foram utilizados nos treinamentos e testes das redes, do mesmo modo que sinais polifônicos. Cada registro disponível foi segmentado em 3 trechos de aproximadamente 0,61 s, com inícios a partir da primeira amostra, da amostra 8001 e da amostra 16001 (ou 12001 no caso dos registros 012PFNO). Assim, foram utilizados 264 sinais de notas simples por piano de ambas as bases.

As combinações foram criadas através da soma dos vetores compostos pelos elementos de cada segmento. Após a soma, cada combinação foi normalizada por sua norma quadrática.

6.6 Treinamento das Redes Neurais

Assim como descrito na Seção 4.6, para cada combinação de notas foram calculadas duas transformadas através do algoritmo rápido da CQT (Seção 2.2). A primeira transformada, para aplicação no primeiro e no segundo grupo de métodos, foi calculada com componentes a partir do *pitch* da nota Lá 0. A segunda transformada, para aplicação apenas no segundo grupo de métodos, foi calculada com componentes a partir do *pitch* da nota mais grave de cada combinação.

Os vetores de entrada das redes neurais foram formados pelos valores absolutos das componentes de cada transformada. Os vetores-objetivo foram formados com 88 elementos, cada um correspondente a uma das notas do piano. A presença de cada nota foi indicada pelo valor 1 no elemento correspondente. As notas ausentes foram indicadas pelo valor 0. As dinâmicas e amostras iniciais escolhidas para cada combinação foram armazenadas para uso na análise dos resultados.

Os pares formados pelos vetores de entrada e vetores-objetivo foram divididos em três conjuntos: um de treino, um de teste e um de validação. O conjunto de treino continha os pares formados a partir de combinações das notas do piano 011PF (011PFNO e 011PFST) e MPP. O conjunto de validação continha os pares formados a partir de combinações das notas do piano 012PF (012PFNO e 012PFST). O conjunto de teste continha os pares formados a partir de combinações das notas do piano 013PF (013PFNO e 013PFST).

Foram realizados testes com um número fixo de vetores nos grupos de treino e validação. Foram criados, para os grupos de treino e validação, 1000 pares de vetores (entrada e objetivo) para cada instrumento, para cada grau de polifonia maior que dois. Além destes, foram utilizados todos os vetores referentes a notas simples e combinações de duas notas. Para o grupo de teste foram criados 2750 pares de vetores para cada instrumento, para cada grau de polifonia maior que dois, e todos os vetores referentes a notas simples e combinações de duas notas.

No total foram realizados testes com conjuntos de treino e validação contendo 37860 pares de vetores (entrada e objetivo). O grupo de teste continha, sempre, 53240 pares de vetores.

Os vetores de entrada dos conjuntos de treino, teste e validação foram escalonados para o uso com redes neurais. Os valores de cada componente foram reduzidos das médias de *ensemble* correspondentes (calculados apenas sobre o conjunto de treino), e divididos pelo dobro dos desvios-padrão de *ensemble* correspondentes (calculados apenas sobre o conjunto de treino).

As redes foram implementadas com as mesmas configurações usadas no treinamento das redes para identificação de notas de violão (descritas na Seção 4.6), utilizando o algoritmo Rprop e o critério de parada descrito na Seção 3.5.

Para encontrar topologias apropriadas para as redes utilizadas em cada método, foram realizados testes com variações na quantidade de neurônios na camada oculta.

Capítulo 7

Implementação e Testes - Piano

7.1 Introdução

Neste capítulo são detalhadas adaptações dos métodos 1A e 2A (desenvolvidas para torná-los compatíveis com a identificação de notas musicais em registros de piano) e os resultados dos testes realizados.

7.2 Método 1A para Piano

Nesta adaptação, os vetores analisados são novamente formados por representações espectrais, obtidas através da CQT, de cada combinação de registros. As notas são classificadas como presentes ou ausentes de acordo com os valores dos elementos do vetor de saída de uma rede neural desenvolvida para o processo de classificação. As notas correspondentes aos elementos com valor maior que 0,5 são classificadas como presentes. Apesar da utilização de combinações com, no máximo, 10 notas simultâneas, durante a execução de um piano é possível obter polifonia igual à quantidade de teclas do instrumento (como descrito na Seção 1.2). Por isto, são aceitos resultados que indiquem mais que 10 notas simultâneas por combinação. Se, para cada vetor de saída, nenhum elemento tiver valor acima de 0,5, apenas a nota correspondente ao maior valor encontrado é classificada como presente na combinação.

Para encontrar configurações apropriadas para as redes foram feitos experimentos com diferentes topologias. As mudanças foram obtidas alterando o número

de neurônios na camada oculta. Foram realizados experimentos utilizando topologias com 406, 381, 356, 331, 306, 281 e 256 neurônios nesta camada. A camada de saída tinha 88 neurônios, o mesmo número de teclas do instrumento.

Na Tabela 7.1 são mostrados os resultados da implementação do método 1A com duas topologias diferentes. A primeira, com tantos neurônios na camada oculta quanto elementos em cada vetor de entrada. A segunda, da qual se obteve o melhor resultado entre as implementações, com 306 neurônios na camada oculta.

Tabela 7.1: Método 1A para Piano

topologia	rede	nº de épocas	acurácia
406 x 88	1	17	24,5
	2	14	22,4
	3	14	23,4

topologia	rede	nº de épocas	acurácia
306 x 88	1	34	**28,5**
	2	16	23,6
	3	20	26,2

As fortes diferenças entre as acurácias das realizações da segunda topologia podem ter sido causadas pela presença de múltiplos mínimos locais na superfície de custo do treinamento. Particularmente, a diferença elevada entre as acurácias obtidas nas duas primeiras realizações (4,9%), bem como a grande diferença no número de épocas de treinamento, indicam problemas de convergência para um mínimo único e global.

Resultados de NER por grau de polifonia são mostrados na Figura 7.1, conjuntamente com resultados apresentados por POLINER e ELLIS [26], obtidos através de seu método de identificação de notas em sinais polifônicos de piano. Poliner e Ellis realizaram suas análises sobre dois bancos de dados, um composto por sinais de piano sintetizados a partir de arquivos MIDI e outro composto por registros de piano automático executados a partir de arquivos MIDI. As medidas apresentadas na Figura 7.1 se referem aos resultados da classificação conjunta dos dois bancos.

Poliner e Ellis apresentaram medidas de NER e de NER em função do grau de polifonia para até oito notas simultâneas. Eles não apresentaram medidas de CER. Além do próprio método, os autores testaram (sobre as mesmas bases de dados) um método proposto por MAROLT [21] para a transcrição de registros de piano e um método proposto por RYYNÄNEN e KLAPURI [24] para transcrição de registros de

instrumentos com *pitch* definido. Como o conjunto de teste desenvolvido para esta dissertação é diferente dos conjuntos analisados por Poliner e Ellis, a comparação dos resultados obtidos da aplicação do método 1A com os resultados obtidos por outros autores pode ser vista apenas como indicativa das vantagens de cada método.

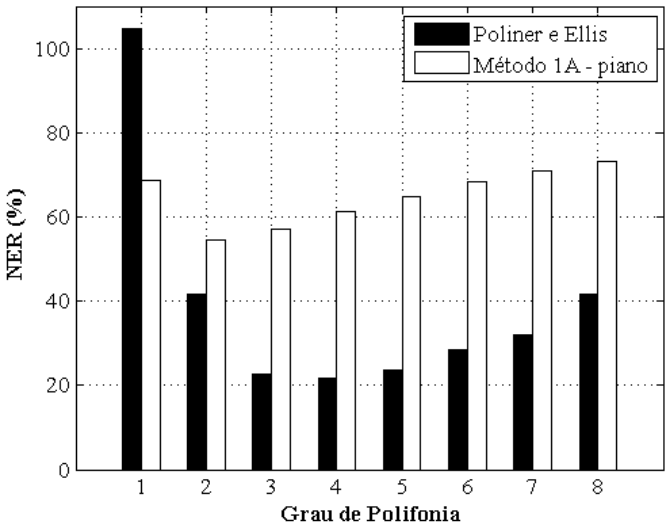

Figura 7.1: Percentuais do NER por grau de polifonia para versões do método 1A referentes aos objetivos 1 e 2.

Na Tabela 7.2 são mostrados os resultados das medições de NER e acurácia obtidos por Poliner e Ellis e os resultados obtidos da aplicação do método 1A.

Tabela 7.2: Resultados de Acurácia e NER para Diferentes Métodos

método	acurácia	NER
1A - piano	28,5	69,7
Poliner e Ellis	67,7	34,2
Ryynänen e Klapuri	46,6	52,3
Marolt	36.9	65.7

7.3 Métodos do Segundo Grupo

Assim como nos métodos do segundo grupo desenvolvidos para a identificação de notas musicais em registros de violão, nos métodos apresentados nesta seção a

identificação de notas de é realizada utilizando duas redes em seqüência. Neles, a primeira rede é utilizada para identificar a nota mais grave de cada combinação e a segunda para encontrar os intervalos entre a nota mais grave de cada combinação e as notas restantes.

As representações espectrais utilizadas no treinamento das redes criadas para identificação da nota mais grave de cada combinação são as mesmas utilizadas para realizar os treinamentos do método 1A para piano.

As representações espectrais utilizadas no treinamento das redes criadas para identificação dos intervalos entre a nota mais grave e as notas restantes são realizadas utilizando os *kernels* criados para o segundo grupo de métodos para identificação de notas de piano, descritos na Seção 6.4. As representações espectrais dos grupos de treino e validação são criadas, para cada combinação, a partir da componente sobre o *pitch* de sua nota mais grave. As representações espectrais do grupo de teste são criadas, para cada combinação, a partir da componente sobre o *pitch* da nota mais grave estimada.

Nas próximas seções são apresentados os métodos desenvolvidos para identificar a nota mais grave de cada combinação e, em seguida, o método desenvolvido para encontrar os intervalos entre a nota mais grave de cada combinação e as notas restantes.

7.3.1 Método 2A para Piano - 1ª etapa

Nesta adaptação da 1ª etapa do método 2A, as redes neurais recebem vetores de entrada formados apenas pelas representações espectrais de cada combinação. Cada vetor-objetivo é formado por 88 elementos, correspondentes, cada um, a uma nota diferente do piano. A presença da nota mais grave é indicada pelo valor 1 no elemento correspondente. Todas os outros elementos do vetor, inclusive os elementos correspondentes a outras notas presentes nas combinações analisadas, recebem o valor 0.

Para cada vetor de saída, a nota correspondente ao elemento com o maior valor é classificada como a nota mais grave da combinação.

Para encontrar configurações apropriadas para as redes foram feitos experimentos com várias topologias. As mudanças foram obtidas alterando-se o número de

neurônios na camada oculta. Foram realizados experimentos utilizando topologias com 431, 406, 381, 356, 331, 306, 281 e 256 neurônios nesta camada. As topologias com, no máximo, tantos neurônios na camada oculta quanto elementos nos vetores de entrada foram testadas primeiro. Como o melhor resultado foi obtido utilizando a topologia com 406 neurônios na camada oculta (mesmo número de elementos dos vetores de entrada), foram realizados novos testes com uma topologia com 431 neurônios na camada oculta. Porém, não foram obtidos resultados melhores com esta configuração.

Na Tabela 7.3 são mostrados os resultados da implementação da 1ª etapa do método 2A para pianos, utilizando redes com 406 neurônios na camada oculta.

Tabela 7.3: Método 2A para Piano - 1ª etapa

rede	nº de épocas	erro nmg
1	87	**45,7**
2	71	47,0
3	69	47,1

Adaptações das primeiras etapas dos métodos 2B e 2C propostos para violão, bem como propostas alternativas para a identificação da nota mais grave, serão desenvolvidas em trabalhos futuros. Neste capítulo, as classificações obtidas na 1ª etapa do método 2A são utilizadas como estimativas para a 2ª etapa.

7.3.2 Método 2A para Piano - 2ª etapa

Assim como o desenvolvimento da 2ª etapa do método 2C para violão independe do desenvolvimento de sua 1ª etapa, o desenvolvimento da 2ª etapa do método 2A para piano também independe de sua 1ª etapa. Novamente, para testar o desempenho do método, foram realizadas avaliações completas e parciais. Na avaliação completa, as transposições dos espectros foram realizadas utilizando as estimativas para as notas mais graves obtidas na 1ª etapa do método. Na avaliação parcial, as transposições dos espectros foram realizadas utilizando, sempre, a informação correta de qual é a nota mais grave de cada combinação (avaliação parcial). Na avaliação completa, o desempenho do método foi medido sobre todas as classifi-

cações obtidas (inclusive as das notas mais graves). Na avaliação parcial, realizada para medir o desempenho da 2ª etapa independentemente dos resultados da 1ª etapa, são descontadas as classificações das notas mais graves.

A criação dos vetores-objetivo para esta etapa é realizada da mesma forma descrita na Seção 5.2.2.4. Os elementos dos vetores-objetivo originais (no caso, vetores usados nos métodos 1A para piano) são deslocados, de modo que o elemento referente à nota mais grave se torne, sempre, o primeiro elemento do vetor.

Foram realizados experimentos utilizando topologias com 431, 406, 381, 356 e 331 neurônios na camada oculta. As topologias com 406, 381, 356 e 331 neurônios na camada oculta foram testadas primeiro. Como o melhor resultado foi obtido utilizando 406 neurônios, foram realizados novos testes utilizando 431 neurônios. Assim como na 1ª etapa do método 2A, os melhores resultados foram obtidos utilizando 406 neurônios na camada oculta.

Na Tabela 7.4 são mostrados os resultados da implementação do método 2A para piano (avaliações parcial e completa), em 3 realizações, na classificação do conjunto de teste, utilizando redes com 406 neurônios na camada oculta.

Tabela 7.4: Método 2A - 2ª etapa - piano

rede	nº de épocas	acurácia (completa)	acurácia (parcial)
1	120	36,8	34,4
2	134	**36,9**	**34,4**
3	126	36,6	34,1

O melhor resultado foi obtido na 2ª realização, com uma rede treinada em 134 épocas. Resultados de NER por grau de polifonia são mostrados na Figura 7.2, conjuntamente com resultados obtidos com a aplicação do método 1A e com os resultados apresentados por Poliner e Ellis.

Exceto pelo resultado para notas simples, a aplicação do método 2A apresentou erros menores que o método 1A. Para nota simples, o aumento do NER foi determinado pelos erros obtidos na 1ª etapa do método 2A somados aos erros de inserção ocorridos na 2ª etapa.

Na Tabela 7.5 são mostrados, além da repetição dos resultados da Tabela 7.2,

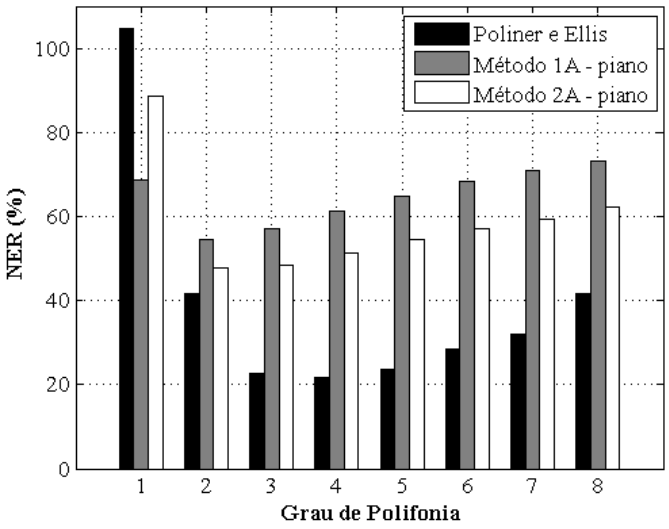

Figura 7.2: Percentuais do NER por grau de polifonia para o método de Poliner e Ellis (na classificação conjunta de sinais sintetizados e de registros reais de piano automático) e para os métodos 1A e 2A (na classificação do conjunto de teste).

os resultados das medições de NER e acurácia das avaliações completa e parcial, obtidas da aplicação do método 2A.

Tabela 7.5: Resultados de Acurácia e NER para Diferentes Métodos

método	acurácia	NER
1A - piano	28,5	69,7
2A - piano (completa)	36,9	59,4
2A - piano (parcial)	34,4	62,9
Poliner e Ellis	67,7	34,2
Ryynänen e Klapuri	46,6	52,3
Marolt	36.9	65.7

Os resultados de acurácia e NER obtidos na avaliação parcial foram piores do que os resultados obtidos na avaliação completa. Isto é, parcialmente, devido à "escala alongada" do piano, discutida na Seção 1.8. Na 2ª etapa do método 2A, as transposições de espectros de diferentes combinações de notas (com os mesmos intervalos) executados em oitavas distantes, não resultam em espectros tão similares entre si quanto as transposições de espectros de combinações (com os mesmos in-

tervalos), executados em oitavas próximas. Isto implica que a rede da 2ª etapa pode receber vetores de entrada muito diferentes, mas que representam intervalos iguais entre a nota mais grave e as notas restantes, reduzindo sua eficiência. Este problema não é tão acentuado na implementação dos métodos do segundo grupo para violão porque este instrumento tem, usualmente, metade da extensão de um piano comum.

7.4 Conclusão

Assim como nos métodos desenvolvidos para a identificação de notas de violão, a divisão do problema em uma etapa de identificação da nota mais grave e outra para identificação dos intervalos entre a nota mais grave e as notas restantes, teve melhores resultados do que a tentativa de estimar todas as notas simultaneamente.

Apesar disto, no método 2A, a etapa de identificação de intervalos entre a nota mais grave e as notas restantes de cada combinação é dificultada pela "escala alongada" do instrumento. Vetores de entrada criados para a 2ª etapa do método, referentes à combinações de notas com os mesmos intervalos, podem não ter máximos nos mesmos elementos, dificultando a identificação dos intervalos.

Capítulo 8

Conclusões

Nesta dissertação foram apresentados métodos desenvolvidos para a identificação de notas musicais em registros de violão solo. Estes métodos têm como base o uso de redes neurais *feed-forward* de múltiplas camadas, treinadas com representações espectrais obtidas através de uma transformada de Q constante. Além destes, também foram apresentadas adaptações voltadas para a identificação de notas musicais em registros de piano.

Os métodos podem ser divididos em duas abordagens: na primeira, apenas uma rede é utilizada na identificação das notas presentes em cada segmento de sinal analisado; na segunda, duas redes são utilizadas em seqüência: a primeira para identificar apenas a nota mais grave de cada segmento de sinal analisado e a segunda para encontrar os intervalos entre a nota mais grave e as notas restantes.

Os métodos criados para identificar notas de violão foram desenvolvidos e aferidos de acordo com uma seqüência de objetivos (designados 1, 2, 3 e 4). De acordo com o objetivo 1 buscou-se identificar notas de combinações formadas a partir de registros com apenas um nível de dinâmica (*mezzo*), extraídos aproximadamente do período de sustentação das notas. De acordo com o objetivo 2 buscou-se identificar notas de combinações similares às desenvolvidas para o objetivo 1, porém, criadas com um registro com dinâmica diferenciada (*forte*) em cada combinação. De acordo com o objetivo 3 buscou-se identificar notas de combinações formadas a partir de registros com a mesma dinâmica (*mezzo*), extraídos, para todas as notas de uma mesma combinação, de um entre três possíveis períodos de segmentação. De acordo com o objetivo 4 buscou-se identificar notas de combinações formadas por registros

que tinham, independentemente, um entre três níveis de dinâmica (*piano*, *mezzo* ou *forte*) extraídos, independentemente, de um entre três períodos de segmentação.

As adaptações dos métodos voltadas para a identificação de notas de piano solo foram desenvolvidas e aferidas buscando-se identificar notas de combinações formadas por registros que tinham, independentemente, um entre três níveis de dinâmica (*piano*, *mezzo* ou *forte*) extraídos, independentemente, de um entre três períodos de segmentação.

Pôde-se observar em experimentos referentes ao objetivo 1 para identificação de notas de violão que o conhecimento prévio do grau de polifonia de cada combinação de notas analisada pode ser utilizado para reduzir o número de ocorrências de erros de inserção, principalmente para combinações com muitas notas. De forma similar, pôde-se observar em experimentos referentes ao objetivo 2 que estabelecer um limite superior para o número de notas que podem ser classificadas como presentes em uma dada combinação contribui para a redução do NER nas classificações de combinações com número de notas próximo ao grau de polifonia do instrumento.

Pôde-se observar em experimentos referentes aos objetivos 2 e 4 que a presença de variações dinâmicas nas combinações analisadas dificultam a identificação das notas que possuem dinâmicas mais baixas. Particularmente nos experimentos referentes ao objetivo 4, onde foram utilizados três níveis diferentes de dinâmica, a maior parte das notas que não foram encontradas tinham dinâmica *piano*. Apenas pequenos percentuais das notas com dinâmica *forte* não foram identificadas.

Pôde-se observar em experimentos referentes ao objetivo 3, onde foram utilizadas redes neurais treinadas com exemplos de combinações de registros com dinâmica *mezzo* extraídos a partir de 3 possibilidades de segmentação (com todos os segmentos de cada combinação extraídos do mesmo período), que a presença de parciais devidas às freqüências de ressonância do violão, principalmente nos períodos de ataque e decaimento, pode provocar muitos casos de falsos positivos, indicando notas que correspondem às faixas de freqüencias onde ocorrem as ressonâncias.

Foi possível observar através da comparação dos resultados obtidos na aplicação dos métodos voltados para objetivo 1 com resultados obtidos na aplicação dos métodos voltados para objetivo 3 que aumentar a variedade de exemplos de treinamento pode aumentar a robustez do processo de classificação. Os resultados

obtidos utilizando os métodos criados para o objetivo 3 na análise de combinações de registros com dinâmica *mezzo* extraídos aproximadamente do período de sustentação foram melhores que os resultados obtidos utilizando os métodos criados para o objetivo 1 na análise do mesmo tipo de combinações.

Para cada objetivo, os resultados dos métodos do segundo grupo, nos quais se divide o problema de identificação de notas musicais em duas etapas (uma para a identificação da nota mais grave e outra para identificação dos intervalos entre a nota mais grave e as notas restantes), foram sempre melhores que os resultados dos métodos do primeiro grupo, nos quais se buscava identificar todas as notas em apenas uma etapa. Isto ocorre porque, dado que a estimativa da nota mais grave esteja correta, a identificação dos intervalos entre a nota mais grave e as notas restantes tem melhor desempenho do que a estimativa direta de todas as notas. Além disto, estimativas para as notas mais graves, obtidas na 1ª etapa dos métodos do segundo grupo, mesmo quando erradas, comumente indicam notas que também pertencem às combinações testadas. Nestes casos, mesmo com a decorrente falha na 2ª etapa, pelo menos uma nota é corretamente indicada.

Para a identificação de notas de piano, os resultados das classificações realizadas em apenas uma etapa não foram promissores. A divisão do problema em uma etapa de identificação da nota mais grave e outra para identificação dos intervalos entre a nota mais grave e as notas restantes, assim como nos métodos desenvolvidos para identificação de notas de violão, produziu melhores resultados.

A etapa de identificação de intervalos entre a nota mais grave e as notas restantes de cada combinação é dificultada pela "escala alongada" do instrumento. Vetores de entrada criados para esta etapa, referentes a combinações de notas com os mesmos intervalos, podem não ter máximos nos mesmos elementos, dificultando a identificação dos intervalos.

Trabalhos Futuros

Devem-se desenvolver propostas alternativas para a etapa de identificação da nota mais grave dos métodos do segundo grupo.

Os métodos para identificação de notas de piano, assim como os métodos

desenvolvidos para violão, devem ser adaptados de acordo com características do instrumento.

Deve ser implementado um método de detecção automática dos períodos do modelo ADSR em função de cada registro analisado.

Novos bancos de dados, compostos por registros de acordes realizados por músicos, em vez de combinações de registros realizadas computacionalmente, devem ser testadas.

Referências Bibliográficas

[1] HERRERA-BOYER, P., KLAPURI, A., DAVY, M., "Automatic Classification of Pitched Musical Instrument Sounds". In: Klapuri, A., Davy, M. (eds.), *Signal Processing Methods for Music Transcription*, New York, USA, Springer, pp. 163–200, 2006.

[2] JÄRVELÄINEN, H., VERMA, T., VÄLIMÄKI, V., "The Effect of Inharmonicity on Pitch in String Instruments Sounds". In: *Proceedings of the International Computer Music Conference*, pp. 237–240, ICMA, Berlin, Germany, August/September 2000.

[3] LEGGE, K. A., FLETCHER, N. H., "Nonlinear Generation of Missing Modes on a Vibrating String", *Journal of the Acoustical Society of America*, v. 76, n. 1, pp. 5–12, July 1984.

[4] FLETCHER, N. H., "The Nonlinear Physics of Musical Instruments", *Reports on Progress in Physics*, v. 62, n. 5, pp. 723–764, May 1999.

[5] EVANS, C., REES, D., "Nonlinear Distortions and Multisine Signals - Part I: Measuring the Best Linear Approximation", *IEEE Transactions on Instrumentation and Measurement*, v. 49, n. 3, pp. 602–609, June 2000.

[6] JENSEN, K., *Timbre Models of Musical Sounds*. Ph.D. thesis, Department of Computer Science, University of Copenhagen, Denmark, July 1999.

[7] WOODHOUSE, J., "Plucked Guitar Transients: Comparison of Measurements and Synthesis", *ACTA Acustica United with Acustica*, v. 90, n. 5, pp. 945–965, September/October 2004.

[8] CHRISTENSEN, O., VISTISEN, B. B., "Simple Model for Low-frequency Guitar Function", *Journal of the Acoustical Society of America*, v. 68, n. 3, pp. 758–766, September 1980.

[9] FIRTH, I. M., "Physics of the Guitar at the Helmholtz and First Top-plate Resonances", *Journal of the Acoustical Society of America*, v. 61, n. 2, pp. 588–593, February 1977.

[10] GOLDEMBERG, R., "Aspectos Acústicos da Afinação de Pianos". In: *Anais do II Seminário de Música, Ciência e Tecnologia*, Unicamp, Campinas, Brasil, Outubro 2005. [online] `http://www.proceedings.scielo.br/scielo.php?script=sci_arttext&pid=MSC0000000102005000100005&lng=en&nrm=iso`.

[11] WARD, W. D., "Musical Perception". In: Tobias, J. V. (ed.), *Foundations of Modern Auditory Theory*, v. 1, New York, USA, Academic Press, pp. 407–459, 1970.

[12] BENSA, J., DAUDET, L., "Efficient Modeling of "Phantom" Partials in Piano Tones". In: *Proceedings of the International Symposium on Musical Acoustics, March 31st to April 3rd 2004 ISMA2004)*, pp. 207–210, ISMA, Nara, Japan, March/April 2004.

[13] NISHIGUCHI, I., "Recent Research on the Acoustics of Pianos", *Acoustical Science and Technology*, v. 25, n. 6, pp. 413–418, November 2004.

[14] CONKLIN JR., H. A., "Piano Strings and 'Phantom' Partials", *Journal of the Acoustical Society of America*, v. 102, n. 1, pp. 659, July 1997.

[15] CONKLIN JR., H. A., "Generation of Partials due to Nonlinear Mixing in a Stringed Instrument", *Journal of the Acoustical Society of America*, v. 105, n. 1, pp. 536–545, January 1999.

[16] NAKAMURA, I., NAGANUMA, D., "Characteristics of Piano Sound Spectra". In: *Proceedings of the Stockholm Music Acoustics Conference, 1993*, pp. 325–330, Stockholm, Sweden, July/August 1993.

[17] CONKLIN JR., H. A., "Design and Tone in the Mechanoacoustic Piano. Part III. Piano Strings and Scale Design", *Journal of the Acoustical Society of America*, v. 100, n. 3, pp. 1286–1298, September 1996.

[18] BONNET, L., LEFEBVRE, R., "High-Resolution Robust Multipitch Analysis of Guitar Chords". In: *114^{th} AES Convention, Preprint 5772*, AES, Amsterdam, The Netherlands, March 2003.

[19] GAGNON, T., LAROUCHE, S., LEFEBVRE, R., "A Neural Network Approach for Pre-classification in Musical Chords Recognition". In: *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, v. 2, pp. 2106–2109, IEEE, November 2003.

[20] SMITH III, J. O., ABEL, J. S., "Bark and ERB Bilinear Transforms", *IEEE Transactions on Speech and Audio Processing*, v. 7, n. 6, pp. 697–708, November 1999.

[21] MAROLT, M., "A Connectionist Approach to Automatic Transcription of Polyphonic Piano Music", *IEEE Transactions on Multimedia*, v. 6, n. 3, pp. 439–449, June 2004.

[22] SZCZUPAK, A. L., BISCAINHO, L. W. P., CALÔBA, L. P., "Identificação de Notas Musicais de Violão Utilizando Redes Neurais". In: *Anais do 4º Congresso de Engenharia de Áudio*, v. 1, pp. 108–112, AES Brasil, São Paulo, Brasil, Maio 2006.

[23] KLAPURI, A., "A Perceptually Motivated Multiple-F0 Estimation Method". In: *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 291–294, IEEE, New Paltz, USA, October 2005.

[24] RYYNÄNEN, M. P., KLAPURI, A., "Polyphonic Music Transcription Using Note Event Modelling". In: *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 319–322, IEEE, New Paltz, USA, October 2005.

[25] RABINER, L. R., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, v. 77, n. 2, pp. 257–286, February 1989.

[26] POLINER, G. E., ELLIS, D. P. W., "A Discriminative Model for Polyphonic Piano Transcription", *EURASIP Journal on Advances in Signal Processing*, v. 2007, January 2007. Article ID 48317, 9 pages.

[27] HAYKIN, S., *Redes Neurais*. 2 ed., Porto Alegre, Brasil, Bookman, 2001.

[28] BROWN, J. C., "Calculation of a Constant Q Spectral Transform", *Journal of the Acoustical Society of America*, v. 89, n. 1, pp. 425–434, January 1991.

[29] DUDA, R. O., HART, P. E., STORK, D. G., *Pattern Classification*. 2 ed., New York, USA, Wiley-Interscience, 2001.

[30] BROWN, J. C., PUCKETTE, M. S., "An Efficient Algorithm for the Calculation of a Constant Q Transform", *Journal of the Acoustical Society of America*, v. 92, n. 5, pp. 2698–2701, November 1992.

[31] DINIZ, P. S. R., SILVA, E. A. B., NETTO, S. L., *Processamento Digital de Sinais: Projeto e Análise de Sistemas*. Porto Alegre, Brasil, Bookman, 2004.

[32] JAIN, A. K., MAO, J., MOHIUDDIN, K. M., "Artificial Neural Networks: A Tutorial", *Computer*, v. 29, n. 3, pp. 31–44, March 1996.

[33] WASSERMAN, P. D., *Neural computing: Theory and Practice*. New York, USA, Van Nostrand Reinhold Co., 1989.

[34] RIEDMILLER, M., BRAUN, H., "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm". In: *Proceedings of the ICNN - International Conference on Neural Networks*, v. 1, pp. 586–591, IEEE, San Francisco, USA, March/April 1993.

[35] TOLLENAERE, T., "Supersab: Fast Adaptive Backpropagation with Good Scaling Properties", *Neural Networks*, v. 3, n. 5, pp. 561–573, October 1990.

[36] JACOBS, R. A., "Increased Rates of Convergence Through Learning Rate Adaptation", *Neural Networks*, v. 1, n. 4, pp. 295–307, December 1988.

[37] GOTO, M., NISHIMURA, T., HASHIGUCHI, H. *et al.*, "RWC Music Database: Music Genre Database and Musical Instrument Sound Database". In: *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pp. 229–230, Baltimore, USA, October 2003.

[38] YEH, C., "RWC Sample Markers Files Including Onset Markers for Instruments in RWC-MDB-I-2001", 2008, `http://recherche.ircam.fr/equipes/analyse-synthese/cyeh/dbfiles/RWC_Markers.zip`.

[39] OPPENHEIM, A. V., WILLSKY, A. S., NAWAB, S. H., *Signals and Systems*. 2 ed., Upper Saddle River, USA, Prentice-Hall, 1997.

[40] HELÉN, M., VIRTANEN, T., "Perceptually Motivated Parametric Representation for Harmonic Sounds for Data Compression Purposes". In: *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx-03)*, London, United Kingdom, September 2003.

[41] RIEDMILLER, M., *Rprop - Description and Implementation Details*, Technical report, Institute für Logik, Komplexität und Deduktionsstyme, University of Karlsruhe, Karlsruhe, Deutschland, January 1994.

[42] KLAPURI, A., VIRTANEN, T., HOLM, J.-M., "Robust Multipitch Estimation for the Analysis and Manipulation of Polyphonic Musical Signals". In: *Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx-00)*, Verona, Italy, December 2000.

[43] DIXON, S., "On the Computer Recognition of Solo Piano Music". In: *Proceedings of the Australasian Computer Music Conference*, pp. 31–37, Brisbane, Australia, July 2000.

[44] OPOLKO, F., WAPNICK, J., "McGill University Master Samples Collection on DVD: Volume 2", DVD, 2006, McGill University.

Apêndice A

Marcações de *Onsets* da Base RWC

A.1 Violões

Tabela A.1: Amostras marcadas como *onsets* na gravação 091CGAFP.

2134	217600	438785	659547	878180	1089240
1307336	1580264	1790968	2012788	2236525	2462827
2688001	2939616	3160769	3388645	3608185	3836013
4050938	4274266	4505197	4718706	4921854	5114365
5331455	5558971	5834351	6058072	6285416	6508141
6737001	6952565	7213565	7427295	7660238	7880431
8104962	8327680	8549895	8800870	9051350	9265883
9499349	9714283	9943044	10164745	10396677	10615300
10838123	11059716	11260316	11488771	11769453	11964538
12216323	12439154	12658179	12908546	13134850	13361670
13588482	13816835	14074369	14253346	14463489	14668289
14875138	15092929	15307780	15531523	15751683	15972352
16215145	16397318	16579071	16750074	16976605	17146486

Tabela A.2: Amostras marcadas como *onsets* na gravação 091CGAFM.

1	193536	402943	619631	840283	1049703
1259215	1479415	1694815	1909755	2129523	2348127
2563073	2783465	2995431	3190366	3406330	3626608
3836020	4056176	4280950	4485224	4704254	4903938
5105263	5310700	5522943	5743215	5958776	6160890
6355971	6569987	6776546	6982661	7247448	7462493
7675514	7885318	8099328	8289272	8504027	8713580
8937166	9152098	9367138	9587712	9826303	10040949
10372098	10595846	10807297	11016706	11255386	11481501
11702010	11911280	12104194	12273663	12483070	12703746
12929535	13144787	13353991	13569130	13789698	13993980
14209026	14429293	14644229	14864899	15074307	15328256
15524972	15693030	15841554	16010056	16191483	16362533

Tabela A.3: Amostras marcadas como *onsets* na gravação 091CGAFF.

585	220258	440939	659458	879718	1100012
1304268	1519192	1715446	1935997	2156654	2371729
2592255	2807912	3009244	3218555	3444837	3670645
3880057	4095088	4321371	4530692	4737030	4918379
5121623	5333504	5583877	5798641	6013566	6238304
6450285	6670851	6885888	7102561	7317601	7524449
7742290	7962228	8175223	8378470	8593506	8780900
9028178	9207391	9433597	9648637	9868800	10089472
10304513	10519554	10734593	10948098	11139688	11338243
11558912	11779484	11988593	12203623	12424193	12639233
12859906	13074844	13295106	13510148	13730817	13980673
14212097	14417921	14629384	14850051	15065089	15285346
15500398	15675909	15839861	16014337	16183395	16338540

Tabela A.4: Amostras marcadas como *onsets* na gravação 092CGAFP.

2016	200927	399558	646809	851618	1059626
1293866	1560257	1815281	2045455	2277630	2506801
2705913	2941775	3162544	3410271	3647996	3874950
4111104	4364677	4559488	4793425	4994712	5201388
5449418	5689506	5915644	6132631	6391842	6592164
6846565	7100769	7310074	7502141	7737656	7942814
8152572	8357193	8582944	8809014	8999665	9223290
9435775	9627357	9864203	10079498	10289823	10479033
10715994	10951607	11146810	11378498	11580820	11742519
11943678	12140762	12332989	12525283	12751596	12976713
13193324	13411893	13654011	13880027	14088803	14334833
14571835	14763644	14967352	15192966	15404036	15638956
15865118	16118616	16327920	16553856	16813124	16985732

Tabela A.5: Amostras marcadas como *onsets* na gravação 092CGAFM.

1395	139776	316896	491200	644816	810696
974632	1147080	1311656	1497352	1671984	1840272
2016496	2196464	2356944	2538864	2708147	2880307
3047955	3223859	3389499	3539643	3701115	3843419
3981691	4140315	4297627	4458491	4618171	4791739
4952539	5114843	5282907	5445243	5593627	5769595
5936699	6114299	6282331	6461915	6617435	6734026
6906010	7038147	7149709	7279581	7380445	7462669
7597197	7761517	7883341	7980589	8101005	8225856
8338432	8434528	8527328	8640768	8786224	8905936
9017456	9155536	9297728	9423008	9510992	9604624
9721520	9825776	9934672	10046016	10209488	10291552
10379568	10467152	10523216	10594240	10670896	10746208

Tabela A.6: Amostras marcadas como *onsets* na gravação 092CGAFF.

2105	226218	485212	733425	962278	1203355
1401951	1644500	1876103	2121370	2332869	2569889
2779245	3004968	3190022	3432606	3650334	3903829
4113227	4301831	4533236	4753743	4985291	5161830
5382338	5554790	5776675	5975030	6184530	6394060
6609054	6807506	7011436	7265006	7518608	7777649
8020311	8257291	8488818	8720978	8947005	9184049
9421085	9669138	9911692	10165221	10413338	10621455
10858609	11095573	11343638	11619261	11884707	12094192
12320204	12535114	12739150	12959648	13180157	13384002
13588058	13797559	13978421	14193403	14419414	14662117
14867417	15033471	15237341	15446859	15676708	15897251
16117653	16342469	16557452	16777953	17042559	17279594

Tabela A.7: Amostras marcadas como *onsets* na gravação 093CGAFP.

2014	187113	396625	606313	815542	1025035
1234546	1401802	1575608	1744418	1917536	2095921
2261680	2440003	2606115	2792227	2972093	3155906
3331853	3508728	3686009	3869400	4066120	4238221
4401644	4566410	4742018	4932291	5136981	5331967
5502519	5685077	5849357	6032614	6216417	6403017
6580905	6764315	6946711	7140106	7295308	7459212
7650755	7812908	7970319	8131421	8322954	8497690
8667661	8873270	9036224	9188972	9347843	9482500
9624995	9788559	9941401	10099941	10277574	10465072
10643732	10815397	10989113	11175122	11343118	11499786
11642319	11804583	11945569	12119385	12286849	12455770
12609610	12812437	12952030	13119765	13275893	13456001

Tabela A.8: Amostras marcadas como *onsets* na gravação 093CGAFM.

1033	210528	442164	662384	888465	1103651
1324609	1540153	1747211	1928418	2124017	2317299
2529445	2734716	2912968	3115114	3302778	3487119
3657404	3845062	4044506	4226307	4414068	4583693
4744022	4919004	5120621	5303442	5499415	5687816
5883383	6055429	6249401	6448620	6643609	6835678
7028891	7206554	7395650	7635451	7833908	8006423
8192902	8377931	8560022	8744344	8938954	9135064
9328094	9532059	9698613	9924685	10106311	10271719
10442616	10608640	10780942	10933301	11109522	11303508
11512984	11695950	11905604	12120658	12298161	12518635
12706017	12893465	13086371	13295876	13485723	13684228
13899202	14117901	14310849	14467782	14682927	14837377

Tabela A.9: Amostras marcadas como *onsets* na gravação 093CGAFF.

1008	155613	291213	478502	649505	831501
1001421	1170205	1340064	1527501	1688905	1830057
1978318	2138902	2332641	2564044	2795520	3010521
3231010	3419617	3651131	3840203	4049702	4225410
4379685	4559491	4753647	4946623	5150603	5327504
5526006	5704507	5897413	6094662	6296808	6494606
6684241	6899622	7120141	7379229	7583204	7825766
8079307	8305417	8536889	8773911	9021959	9248070
9474073	9755133	9981136	10254397	10530011	10723004
10937983	11158402	11378970	11571928	11781482	11985304
12183818	12361101	12570644	12769163	12960902	13214527
13399491	13597902	13771021	13980504	14189952	14399307
14619908	14839566	15054557	15226294	15452309	15661718

A.2 Pianos

Tabela A.10: Amostras marcadas como *onsets* na gravação 011PFNOF.

2849	213784	428312	643096	852248	1061144	1272344
1475864	1681176	1889304	2094104	2298904	2503704	2712088
2922008	3125016	3325464	3533080	3739928	3940376	4143128
4344600	4544792	4742680	4945176	5148184	5346072	5544984
5747480	5942040	6139928	6335768	6533144	6732824	6928664
7128344	7324184	7522072	7717912	7910936	8106520	8304664
8500760	8701208	8901400	9100568	9292824	9488920	9684504
9884952	10085656	10284568	10481176	10675992	10876696	11067672
11272059	11462267	11654267	11844219	12033915	12216443	12406139
12593019	12782971	12967291	13155963	13344960	13532283	13721723
13918587	14112379	14313083	14507643	14711421	14913403	15107195
15306619	15503739	15697283	15894907	16092283	16286075	16480635
16675707	16865403	17055867	17247355			

Tabela A.11: Amostras marcadas como *onsets* na gravação 011PFNOM.

4535	202240	412017	630920	846079	1064021	1278921
1489919	1703603	1913445	2123751	2326629	2535297	2745099
2950980	3161091	3360186	3564299	3767147	3969781	4173268
4372069	4573285	4780320	4980922	5182373	5382146	5580901
5782117	5986621	6189472	6389557	6591143	6794301	6995141
7201891	7404171	7608081	7809741	8010354	8211112	8405605
8608436	8807343	9005793	9200741	9406933	9600101	9796197
9971813	10165349	10367362	10563911	10723468	10903141	11102274
11293797	11489893	11688920	11882704	12050021	12204943	12399169
12591783	12762213	12951969	13141048	13333070	13521611	13742933
13966076	14203083	4449785	14638580	14866136	15106240	15323714
15483083	15662546	15857456	16063155	16265168	16474611	16672017
16867322	17064707	17257472	17451609			

Tabela A.12: Amostras marcadas como *onsets* na gravação 011PFNOP.

3267	156528	319758	485099	666715	844759	1017808
1189329	1378213	1542434	1732473	1911662	2083072	2260882
2434339	2601425	2801171	2981661	3169466	3354513	3550451
3715972	3900623	4087342	4255728	4421378	4598935	4793654
5003807	5229923	5420513	5634617	5833403	6037764	6228292
6406285	6593543	6788288	6988810	7200279	7418777	7606374
7806160	8008933	8202378	8401830	8610022	8807146	9006932
9204892	9405671	9570560	9768431	9959234	10144642	10335269
10541548	10754344	10962967	11171129	11375014	11566713	11732689
11900800	12070260	12241659	12428052	12598200	12756504	12930054
13130127	13391408	13602367	13837763	14014163	14190563	14366963
14543363	14719763	14896163	15072563	15248963	15426752	15585662
15778163	15955720	16130963	16247120			

Tabela A.13: Amostras marcadas como *onsets* na gravação 011PFSTF.

12784	132861	237071	346140	457310	563644	672499
783352	888647	993523	1096710	1204822	1306137	1415251
1514876	1612698	1711095	1799153	1892241	1978348	2070992
2158783	2248834	2337432	2429340	2518939	2605033	2687000
2765931	2850073	2937288	3022651	3105601	3192880	3273103
3355925	3439013	3516684	3597476	3680969	3769241	3852203
3932677	4011195	4091163	4180177	4271285	4361493	4449294
4532655	4609793	4690915	4783657	4867687	4946230	5025144
5107104	5197260	5274247	5360650	5445011	5522941	5607976
5690396	5768405	5842673	5920424	5999847	6075078	6163585
6284820	6391585	6503848	6600467	6697288	6799078	6891786
6977662	7062904	7146771	7222795	7302759	7388790	7471334
7549750	7617980	7691268	7767606			

Tabela A.14: Amostras marcadas como *onsets* na gravação 011PFSTM.

17844	147292	281000	417238	554920	687996	794858
914133	1036194	1159318	1288424	1406992	1545251	1658731
1767122	1889589	2010807	2136804	2257446	2371759	2499630
2623994	2736274	2857087	2978780	3097499	3219484	3345985
3472511	3601322	3725913	3845317	3965231	4083310	4190441
4307415	4417915	4528729	4641325	4771286	4893178	5024072
5150024	5272614	5403644	5528497	5661044	5785490	5911245
6039857	6166960	6292377	6428189	6560468	6695705	6834213
6971278	7105381	7237656	7375916	7509539	7647400	7780154
7910734	8052046	8188766	8322907	8458554	8595261	8746847
8953428	9133782	9328598	9521809	9714535	9885349	10049647
10211169	10365799	10515898	10684383	10834367	10994089	11158391
11309761	11458519	11605354	11762496			

Tabela A.15: Amostras marcadas como *onsets* na gravação 011PFSTP.

15352	153082	304614	479356	654186	818739	988416
1139004	1282836	1440435	1604252	1779316	1954011	2121694
2299846	2469520	2615297	2768648	2920038	3071889	3217500
3377610	3536022	3691864	3845828	4015528	4191264	4346238
4503808	4664086	4817298	4948375	5113056	5271168	5429409
5590648	5750649	5911418	6074798	6256914	6439354	6593461
6751207	6908338	7062887	7215174	7383213	7573180	7755447
7911771	8068157	8229836	8427658	8582052	8732479	8883747
9040660	9196862	9387788	9612254	9815236	9976293	10134712
10300116	10457636	10625091	10774396	10929381	11083597	11272969
11535417	11826587	12051320	12269710	12519551	12764732	12929538
13107450	13276805	13434060	13623644	13775119	13933865	14089762
14260681	14440266	14615756	14797134			

Tabela A.16: Amostras marcadas como *onsets* na gravação 012PFNOF.

11330	150245	297187	449332	603126	736749	879232
1013479	1148544	1278916	1401984	1535616	1667342	1798184
1933010	2063616	2193152	2333314	2469149	2594712	2727012
2867072	2999640	3137232	3264240	3399186	3529610	3667751
3803107	3929600	4056971	4194596	4335406	4475520	4613201
4755200	4903839	5055508	5187328	5320320	5461938	5597885
5747712	5881856	6016754	6142439	6264155	6378112	6505728
6613760	6734848	6869952	7004736	7131328	7242048	7367936
7460736	7541568	7624192	7736320	7840256	7971072	8087040
8184064	8263936	8344280	8424192	8506384	8601792	8693504
8783021	8881554	8970848	9068928	9161440	9250016	9341303
9418354	9497088	9566080	9634560	9704576	9775488	9843072
9906048	9972608	10034432	10104576			

Tabela A.17: Amostras marcadas como *onsets* na gravação 012PFNOM.

4288	164816	337248	504373	673792	847275	1016192
1192834	1355712	1533239	1704480	1876288	2047963	2211264
2366464	2530560	2689284	2849802	3028224	3194624	3369984
3533763	3697408	3844630	3998784	4155474	4312087	4473856
4631552	4784768	4936064	5105417	5246342	5388160	5535380
5677568	5816064	5930916	6080768	6249984	6375424	6479872
6605539	6720028	6828937	6919840	7030013	7126229	7216352
7304232	7403392	7545984	7668352	7764569	7864576	7958784
8045696	8126848	8203904	8302720	8397824	8511872	8608640
8685952	8756672	8828928	8900672	8969152	9040960	9115840
9186752	9250240	9316160	9382025	9442432	9536447	9616540
9691244	9770243	9842702	9914135	9991624	10066960	10178181
10279639	10404416	10511867	10598012			

Tabela A.18: Amostras marcadas como *onsets* na gravação 012PFNOP.

9920	202448	400015	596126	779191	967360	1153600
1343348	1524135	1700368	1875239	2053600	2241408	2418208
2587290	2766560	2948539	3131751	3329176	3512551	3687187
3853918	4005374	4168826	4329054	4503138	4667744	4829906
5022304	5211146	5382720	5551625	5729259	5906225	6074624
6247334	6429723	6566208	6749696	6924960	7077566	7199786
7334732	7474123	7604244	7719040	7831436	7964032	8091083
8200192	8319962	8444080	8592512	8708800	8871152	8994963
9169975	9349056	9452904	9562952	9679694	9789572	9900128
9997825	10111972	10221708	10318086	10421632	10535704	10655136
10754916	10860333	10966352	11087208	11188272	11287527	11386728
11508776	11628472	11738202	11865492	12003822	12138689	12249856
12377696	12500704	12636816	12751704			

Tabela A.19: Amostras marcadas como *onsets* na gravação 012PFSTF.

18236	165128	293588	447518	569728	687888	812752
932960	1055680	1178464	1302784	1423616	1544096	1648032
1762053	1873592	1984256	2102071	2217212	2327245	2430800
2550324	2666772	2778052	2879840	2982080	3086944	3193632
3290656	3396784	3508576	3612135	3732416	3843632	3953904
4065840	4169728	4276080	4378736	4485488	4587694	4690256
4801616	4901568	5021840	5119408	5223072	5318112	5420240
5523024	5626800	5733520	5855504	5951168	6056144	6156192
6252544	6355731	6463824	6560616	6674464	6778240	6889984
6995872	7117408	7218528	7317280	7427286	7515444	7625312
7722976	7860992	7966064	8088768	8202954	8319744	8442864
8553872	8669984	8803771	8918929	9029048	9137220	9263442
9373184	9463598	9567109	9675253			

Tabela A.20: Amostras marcadas como *onsets* na gravação 012PFSTM.

12032	153792	282880	424576	547136	682496	800768
908580	1039731	1163487	1296864	1414656	1533280	1671872
1783648	1903136	2030880	2162400	2294048	2438353	2542848
2662603	2778126	2905561	3003649	3141320	3262577	3372107
3483337	3587922	3722671	3822456	3937814	4082029	4193842
4327412	4461274	4577664	4676544	4799424	4917275	5046130
5176349	5293096	5390070	5501597	5617920	5727111	5829008
5946582	6043920	6156608	6252160	6371488	6475584	6577136
6694672	6806896	6907952	7032736	7142368	7262336	7372602
7474592	7609255	7754828	7860908	7972268	8102336	8212072
8328288	8455810	8572888	8697080	8800472	8919504	9034080
9169832	9300840	9408688	9525120	9633320	9747120	9846392
9966008	10062424	10171848	10285088			

Tabela A.21: Amostras marcadas como *onsets* na gravação 012PFSTP.

9376	164640	314144	456416	596832	736896	866464
988160	1124028	1252576	1394176	1515908	1638304	1761856
1895840	2037696	2168682	2294748	2425799	2543815	2648399
2765772	2896039	3019875	3143152	3252719	3373568	3491541
3604848	3721332	3841297	3959824	4079376	4223070	4350144
4458976	4596608	4715936	4840467	4961664	5078265	5200916
5325953	5449536	5559817	5673595	5797216	5916544	6049440
6178318	6307200	6448495	6552752	6670875	6799269	6924519
7053328	7165456	7287680	7419236	7533856	7658176	7782192
7911904	8031856	8148104	8267264	8390576	8521107	8646221
8759156	8897936	9050572	9195160	9333968	9464568	9595864
9704880	9840712	10021826	10169808	10336720	10503112	10656192
10829720	10981325	11135297	11282552			

Tabela A.22: Amostras marcadas como *onsets* na gravação 013PFNOF.

23526	213812	397204	594869	791418	977345	1181496
1376836	1582549	1778079	1967712	2159708	2366773	2572815
2773567	2972627	3186291	3407200	3624960	3816288	3998746
4191486	4394496	4602840	4810432	5012928	5213024	5397024
5597824	5790752	5981240	6165729	6352656	6544838	6719290
6929118	7156352	7365338	7587672	7785280	8001408	8195744
8392437	8579185	8761940	8939218	9112160	9296217	9463104
9658522	9845408	10020160	10192272	10380470	10552837	10726400
10912374	11102832	11251551	11433609	11593088	11768735	11924464
12057888	12232480	12386512	12517296	12634144	12764677	12874980
12994790	13119864	13234944	13359640	13478896	13597944	13706440
13806376	13909712	14048271	14171464	14288693	14380240	14474210
14565841	14673616	14767348	14864080			

Tabela A.23: Amostras marcadas como *onsets* na gravação 013PFNOM.

21027	196415	378560	570918	780544	976672	1162720
1350656	1568928	1780864	1977568	2180608	2381312	2577312
2776672	2983296	3180672	3386688	3599616	3797984	4006656
4220483	4424416	4629280	4832400	5020702	5193701	5373632
5557568	5751326	5935168	6113888	6313184	6526115	6723232
6920672	7114720	7296896	7490528	7687392	7887424	8074368
8266752	8458976	8648288	8838208	9001984	9195648	9385184
9559269	9743213	9931373	10104992	10284000	10469344	10656122
10836093	11020288	11199363	11386251	11568032	11757696	11937376
12123277	12294880	12462528	12638592	12804688	12981904	13143356
13311808	13526285	13732432	13960846	14194784	14389403	14610710
14822487	15044850	15228991	15432087	15641485	15857600	16057612
16243587	16437117	16621075	16829381			

Tabela A.24: Amostras marcadas como *onsets* na gravação 013PFNOP.

24178	218348	443729	667980	884408	1099457	1314859
1539116	1783306	2017790	2247601	2467786	2741324	2971354
3195165	3416911	3649149	3850050	4077733	4299522	4525545
4731729	4945913	5154110	5372294	5584108	5815633	6041074
6271822	6486290	6711232	6932416	7149610	7351743	7560059
7766976	7976838	8184146	8395040	8620178	8843029	9072279
9294114	9495056	9719372	9924966	10132879	10360435	10510656
10707968	10911328	11119808	11325433	11542336	11732096	11994656
12196928	12413120	12614336	12820256	13031968	13233536	13431360
13575468	13794538	14036416	14252064	14438816	14637344	14841963
15065376	15258272	15487801	15688576	15919488	16129726	16355721
16581720	16789856	16999061	17219957	17439808	17637836	17833899
18016128	18203648	18409529	18570063			

Tabela A.25: Amostras marcadas como *onsets* na gravação 013PFSTF.

23780	158682	297480	446059	590588	727829	874890
1015982	1158117	1312752	1463062	1603921	1746110	1889985
2054214	2188168	2342360	2486306	2641230	2781527	2930678
3088918	3227907	3373392	3519850	3687727	3825423	3971928
4127791	4265396	4402642	4526471	4656661	4805037	4958022
5110096	5265396	5416013	5552317	5708314	5828425	5961010
6114649	6276362	6437122	6567114	6714183	6860402	7006436
7160904	7311387	7451894	7598961	7735222	7892215	8030637
8165826	8306210	8438226	8567169	8684983	8817532	8963307
9117159	9270822	9394732	9537336	9676845	9829368	9963043
10108209	10250861	10397582	10538211	10691433	10862465	11000856
11159251	11322199	11492510	11647936	11807199	11950027	12099791
12242899	12412059	12559282	12712427			

Tabela A.26: Amostras marcadas como *onsets* na gravação 013PFSTM.

25838	137276	253408	381380	508771	630808	739432
848924	972928	1087686	1214848	1325974	1446207	1577740
1701740	1834912	1955452	2072234	2202465	2335256	2457514
2573374	2684618	2806032	2931508	3045867	3160420	3288386
3405503	3512042	3639300	3759356	3879802	4002046	4120307
4237540	4366797	4500555	4627672	4739136	4854178	4963746
5071281	5190441	5306785	5418842	5537326	5649848	5758520
5875994	5989468	6107085	6215040	6329636	6440141	6554214
6666161	6793452	6911374	7032974	7147783	7266594	7380250
7510793	7638721	7745156	7856606	7973053	8099737	8221050
8330235	8457523	8581385	8729722	8877076	9025841	9196440
9347916	9484921	9622185	9764358	9893179	10033656	10177933
10306913	10456356	10588914	10746216			

Tabela A.27: Amostras marcadas como *onsets* na gravação 013PFSTP.

25632	152641	281580	410748	534526	666784	784014
906932	1035890	1170692	1288402	1417250	1545511	1689218
1832330	1955072	2080010	2198268	2321609	2444422	2564116
2686877	2796728	2926106	3062912	3193596	3319860	3439503
3555820	3680203	3793690	3918346	4053200	4179916	4294544
4409388	4559816	4677876	4784208	4890068	5037004	5178788
5311786	5450760	5582216	5701644	5820322	5945064	6068102
6199579	6328050	6446272	6567279	6697456	6866390	7011822
7153957	7289587	7423998	7550972	7686714	7828360	7991228
8147468	8311466	8468986	8624477	8771851	8924392	9070246
9200764	9352740	9502803	9650706	9802067	9954055	10130704
10275205	10433288	10571686	10720657	10869809	11024785	11177489
11318024	11472633	11614351	11765708			