Universidade Federal do Rio de Janeiro

Escola Politécnica

Departamento de Eletrônica e de Computação

Algoritmos para Reconstrução da Fase de Sinais de Áudio

Autor:

Carlos Vinícius Caldas Campos

Orientador:

Luiz Wagner Pereira Biscainho, D.Sc.

Coorientador:

Alexandre Leizor Szczupak, M.Sc.

Examinador:

Alan Freihof Tygel, M.Sc.

Examinador:

José Gabriel Rodríguez Carneiro Gomes, Ph.D.

1 de março de 2011

DEL

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
Escola Politécnica - Departamento de Eletrônica e de Computação
Centro de Tecnologia, bloco H, sala H-217, Cidade Universitária
Rio de Janeiro - RJ CEP 21949-900

DEDICATÓRIA

Dedico este trabalho à minha família, aos meus amigos e à minha namorada.

AGRADECIMENTO

Gostaria de agradecer à minha família pelo carinho e suporte durante a minha formação. Aos meus amigos em geral, agradeço pela agradável companhia durante essa caminhada. Para alguns destes, gostaria de elaborar agradecimentos pontuais.

Agradeço a Bernardo Aquino pelas discussões de conhecimentos e integrais, cena comum no início da faculdade, tempo em que as outras pessoas da nossa turma apenas desconfiavam que fossemos loucos.

Agradeço a Daniel Cayres por várias conversas no ônibus e garrafas de mate que bebemos juntos para ajudar a aguentar o tranco dos trabalhos intermináveis.

Agradeço a Dayana Lole por outras tantas conversas. Conversas que podiam ser longas ou curtas, mas que sempre tinham suas partes reflexivas, divertidas, pontos de discordância, concordância. Enfim, agradeço por todo o jeito, simpatia e estresse de vez em quando.

Agradeço a Felipe Clark por compartilhar sua visão bem-humorada e crítica das coisas do nosso cotidiano e por várias discussões teóricas bastante produtivas.

Agradeço a Renan Mariano, minha dupla em inúmeros trabalhos, tanto por sua contribuição fundamental no início desse trabalho quanto por me aguentar reclamando de tudo e continuar rindo e criando suas músicas e piadas. Bem, isso fora as piadas que fazemos para ele.

Gostaria de agradecer especialmente a Isabela Apolinário, minha namorada, por toda a força, motivação e felicidade que a sua chegada à minha vida trouxe.

Por fim, meus agradecimentos aos membros da banca examinadora. Agradeço aos examinadores, José Gabriel e Alan Tygel, pelo pronto aceite do convite para participar da banca. Aos meus orientadores, Alexandre Leizor e Luiz Wagner, agradeço pela enorme paciência e dedicação de seu tempo, o que tornou possível o desenvolvimento deste trabalho.

RESUMO

Este trabalho trata da descrição, implementação e testes de quatro métodos iterativos encontrados na literatura para reconstrução da fase de sinais de áudio a partir de seus espectrogramas: Griffin e Lim (G&L), Real-Time Iterative Spectrogram Inversion (RTISI), Real-Time Iterative Spectrogram Inversion with Look Ahead (RTISI-LA) e Multiple-Input Spectrogram Inversion (MISI).

São apresentados testes de cada método para diferentes tipos de sinais, tendo seus resultados criticamente comparados de forma a determinar parâmetros mais recomendados para sua utilização.

Dentre os métodos G&L, RTISI e RTISI-LA, que necessitam apenas do espectro de magnitude do sinal, o método RTISI-LA apresentou melhor desempenho nos testes realizados.

Já o método MISI, aplicável apenas em situações nas quais um sinal de mistura também esteja disponível, apresentou, em geral, melhor desempenho que os outros métodos testados. No entanto, houve casos em que os métodos RTISI e RTISI-LA tiveram melhor desempenho.

Foram propostas ainda algumas modificações referentes aos critérios de parada dos métodos. Estas modificações constituem mudança no funcionamento geral apenas para os métodos RTISI e RTISI-LA, não tendo se mostrado vantajosas para estes métodos.

Palavras-Chave: reconstrução de fase, estimação de sinais, inversão de espectrogramas.

ABSTRACT

This work presents the general description, implementation and tests of four iterative methods found in the literature for reconstructing the phase of audio signals from their spectrograms: Griffin and Lim (G&L), Real-Time Iterative Spectrogram Inversion (RTISI), Real-Time Iterative Spectrogram Inversion with Look Ahead (RTISI-LA) and Multiple-Input Spectrogram Inversion (MISI).

Tests of each method using different kinds of signals are presented and their results are compared in order to determine a set of appropriate parameters for their usage.

Among G&L, RTISI and RTISI-LA methods, which require only the availability of the signal magnitude spectrogram, the latter yielded the best results in the tests performed.

The MISI method, which is applicable only when a mixture signal is also available, presented, in general, better performance than the other methods — even though, for some tests, the RTISI and RTISI-LA methods exhibited better results than the MISI.

Some modifications in the methods stopping criteria were proposed. These modifications constitute some operational changes just for the RTISI and RTISI-LA methods, and brought no improvements to their performances.

Keywords: phase reconstruction, signal estimation, spectrogram inversion, phase retrieval.

Sumário

Lista de Figuras

Lista de Tabelas

SIGLAS

G&L - Griffin e Lim

RTISI - *Real-Time Iterative Spectrogram Inversion*

RTISI-LA - *Real-Time Iterative Spectrogram Inversion with Look Ahead*

MISI - *Multiple-Input Spectrogram Inversion*

STFT - *Short-Time Fourier Transform*

MSTFT - *Modified Short-Time Fourier Transform*

DFT - *Discrete Fourier Transform*

FFT - *Fast Fourier Transform*

IFFT - *Inverse Fast Fourier Transform*

SiSEC - *Signal Separation Evaluation Campaign*

Capítulo 1

Introdução

1.1 Tema

Este trabalho se insere na área de processamento digital de sinais de áudio, tratando em particular do problema da reconstrução de espectros de fase a partir de espectros de magnitude referentes a quadros sequenciais de um sinal.

1.2 Delimitação

O projeto abrange o estudo de técnicas de reconstrução de fase aplicadas a espectrogramas obtidos a partir de sinais sonoros, em particular tonais e percussivos. Será avaliado o desempenho dessas técnicas, tomando por base métricas objetivas sugeridas na literatura.

1.3 Justificativa

No contexto dos trabalhos de separação de fontes, comumente deve-se estimar a fase de sinais a partir de espectros de magnitude [1, 2]. Estes espectros são as informações disponíveis de cada fonte. Nas aplicações que requerem que os sinais de cada fonte separada sejam ressintetizados, pode não ser adequado utilizar a fase da mistura para reconstruí-los. A associação da fase da mistura com os espectros de magnitude obtidos no processo de separação pode criar defeitos e distorções no sinal resultante que impossibilitem obter uma qualidade aceitável. Por isto,

buscou-se investigar algoritmos de reconstrução de fase, implementados com base nas descrições da literatura [3, 4, 5]. Isto é feito para sinais sonoros disponibilizados pelos autores ou sinais com características semelhantes àqueles citados por eles, avaliando o desempenho dos métodos de forma comparativa. Adicionalmente, são sugeridas melhorias para alguns algoritmos.

1.4 Objetivos

Este trabalho visa a fazer uma comparação sistemática do desempenho de alguns métodos de reconstrução de fase da literatura e determinar parâmetros mais apropriados para a sua utilização. Esta comparação é importante para que se consiga escolher o melhor algoritmo a ser utilizado de acordo, por exemplo, com o tipo de sinal que se deseja sintetizar. Discussões intuitivas quanto à quantidade de operações necessárias para cada técnica são realizadas, sem, no entanto, apresentar análises detalhadas de custo computacional.

1.5 Histórico

Por volta dos anos 80, havia a preocupação com o desenvolvimento de sistemas que fossem capazes de gerar sinais de boa qualidade a partir de espectros de potência. Esta preocupação tinha base em aplicações como a mudança de escala de tempo de sinais e a redução de ruído por subtração espectral [6, 7].

O algoritmo de Griffin e Lim (G&L) [3] foi desenvolvido como uma das primeiras soluções para o problema. Ele adota a abordagem de minimização de erros quadráticos médios entre o espectro do sinal estimado e o espectro dado.

Essa área, por vezes referida como *phase retrieval*, teve sua importância aumentada com a difusão de técnicas de separação de fontes que realizam os cálculos apenas sobre a magnitude do espectro [1, 2]. Visando à redução da complexidade computacional, foram desenvolvidos os algoritmos de Inversão Iterativa de Espectrograma em Tempo Real (RTISI, do inglês *Real-Time Iterative Spectrogram Inversion*) [4, 8], voltados para aplicações de tempo real. A diferença consistiu em procurar reconstruir o sinal quadro a quadro, em vez de reconstruí-lo como um todo, como no algoritmo G&L.

Com o enfoque voltado ao desenvolvimento da área de Separação Cega de Fontes, mais recentemente foi publicado o método de Inversão de Espectrograma de Múltiplas Entradas (MISI, do inglês *Multiple-Input Spectrogram Inversion*) [5], que, apesar de utilizar um processo análogo àquele proposto por Griffin e Lim [3], inovou ao utilizar um sistema de malha fechada para a estimação de fase. Neste método, o sinal de mistura é utilizado para a verificação da coerência da estimativa de fase de todos os sinais envolvidos. O sinal da mistura é subtraído da sobreposição dos sinais estimados para as fontes, formando um sinal de erro que, realimentado, influencia a próxima estimativa dos sinais das fontes.

1.6 Metodologia

A primeira etapa do trabalho consistiu em levantar as bases de dados indicadas na literatura da área e escolher sinais adicionais de forma a permitir a realização de testes para os métodos de reconstrução de fase escolhidos.

Os métodos tiveram sua base teórica estudada e foram implementados segundo a descrição dos artigos nos quais foram apresentados.

Os sinais das bases de dados foram submetidos aos métodos selecionados e a qualidade dos resultados foi aferida através de comparações com sinais de referência. Para tanto, foram usadas medidas objetivas [3, 4] propostas nos artigos dos métodos.

Os métodos escolhidos foram os algoritmos G&L [3], duas versões do RTISI (RTISI e RTISI-LA) [4, 8], e o MISI [5].

1.7 Estrutura do Texto

O Capítulo 2 discorre sobre os métodos de reconstrução, explicitando seus desenvolvimentos teóricos, bem como considerações adotadas em suas implementações computacionais. A dedução das principais equações utilizadas nos métodos é explicitada, e seus significados são discutidos intuitivamente.

A metodologia e os resultados dos testes propostos são mostrados no Capítulo 3. Estes resultados são analisados, traçando-se comparações com os aspectos esperados.

No Capítulo 4 são apresentadas as conclusões quanto ao funcionamento dos métodos como um todo, explicitando aplicações mais apropriadas, apontadas pelos testes. São mostradas possibilidades de trabalhos futuros, abrangendo partes deste trabalho que podem vir a demandar estudos mais aprofundados.

Capítulo 2

Métodos de Reconstrução

Em aplicações como a separação de fontes sonoras e a mudança da escala temporal de sinais de áudio, há técnicas que lidam apenas com espectros de magnitude [1, 2]. Estas técnicas desconsideram espectros de fase, faltando, assim, informação suficiente para a geração de sinais temporais coerentes com os espectrogramas disponíveis.

Neste capítulo são descritos métodos para a estimação dos espectros de fase que possibilitam gerar sinais temporais a partir de espectrogramas.

2.1 Terminologia

O algoritmos de reconstrução de fase buscam sintetizar um sinal temporal x coerente com um espectrograma dado. Uma solução adotada é a utilização de métodos iterativos. Neste trabalho, foram escolhidos os métodos G&L, RTISI, RTISI-LA e MISI. Os desenvolvimentos das fórmulas de atualização e das bases teóricas [3, 4, 5, 8] referentes a cada método são mostrados nas Seções 2.2, 2.3, 2.4 e 2.5. Nesta seção são apresentados terminologia e conceitos básicos para a realização desses desenvolvimentos.

No processamento digital de sinais de áudio, usualmente os sinais são divididos em sequências de segmentos de pequena duração, chamados quadros. Para cada quadro, pode-se considerar, com boa precisão, que o sinal é estacionário, tornando possível a aplicação de diversas ferramentas desenvolvidas para esta família de sinais.

A obtenção desses segmentos pode ser modelada matematicamente através da multiplicação do sinal original por uma função de comprimento finito. Essa função é chamada de janela e será representada por w ao longo deste trabalho. O processo de aplicação de janelas ao sinal, chamado janelamento, pode ser descrito como

$$y_w[mS, l] = \sum_{l=-\infty}^{\infty} y[l]w[l - mS], \tag{2.1}$$

onde S é chamado de passo (em amostras) entre janelas consecutivas e m é o índice de um quadro. É importante notar que qualquer análise de y realizada com o auxílio de y_w será influenciada pela escolha da janela. O índice w é, então, utilizado para enfatizar essa influência.

Muitas vezes é de grande utilidade caracterizar a sequência de quadros através de uma representação tempo-frequência. Através dela pode-se, por exemplo, visualizar a variação do espectro de magnitude do sinal ao longo do tempo. Uma ferramenta frequentemente utilizada para essa representação é a Transformada de Fourier de Curta Duração (STFT, na sigla em inglês).

A STFT de um sinal y, definida com o auxílio da Equação 2.1, é dada por

$$Y_w(mS, \omega) = \sum_{l=-\infty}^{\infty} y_w[mS, l]e^{-j\omega l}, \tag{2.2}$$

onde ω é a frequência angular.

A partir da STFT, podem ser definidas representações auxiliares para o sinal. Uma delas é o espectrograma de potência, que é dado por

$$E(mS, \omega) = |Y_w(mS, \omega)|^2 . \tag{2.3}$$

Com base nas Fórmulas (2.3) e (2.2), vê-se que o espectro de magnitude da STFT de um sinal, também chamado de espectrograma de magnitude, pode ser facilmente obtido a partir do espectrograma de potência. Para isto, basta que se extraia a raiz quadrada dos valores deste espectrograma. Porém, dado um espectrograma, o espectro de magnitude $|Y_w(mS, \omega)|$ inferido pode não corresponder necessariamente à magnitude de uma STFT possível para um sinal [3]. Isto pode ser devido à modificações realizadas diretamente sobre Y_w, em vez de realizadas sobre o sinal temporal y, ou ao fato de o próprio espectro de magnitude ser resultado de um processo de estimação. Portanto, este espectro de magnitude é referido como

sendo de uma Transformada de Fourier de Curta Duração Modificada (MSTFT, na sigla em inglês), de forma a ressaltar sua diferença com relação ao de uma STFT.

Deve-se notar que, dado o espectrograma, não haverá uma informação de fase associada, como dito anteriormente. Assim, a resolução do problema com a obtenção do sinal real pode ser vista como tendo dois objetivos: a reconstrução da fase do sinal e a obtenção de um espectro de magnitude realizável o mais próximo possível ao da MSTFT obtida a partir do espectrograma.

Em particular, o conceito da MSTFT é bastante importante neste trabalho, sendo frequentemente citado nos desenvolvimentos apresentados nas seções seguintes.

2.2 G&L (*Griffin e Lim*)

2.2.1 Desenvolvimento Teórico

Utilizando a MSTFT $|Y_w(mS,\omega)|$, dada, e a janela, o algoritmo G&L realiza iterativamente o cálculo de um sinal temporal x que a cada iteração tem o espectro de magnitude de sua STFT mais próximo de $|Y_w(mS,\omega)|$[3]. Esta aproximação é realizada através da minimização de um erro quadrático médio. Para isto, utiliza-se uma fórmula de atualização que pode ser deduzida da seguinte forma:

Sejam Y_w uma MSTFT, obtida pela associação do espectro de magnitude $|Y_w(mS,\omega)|$ a um espectro de fase identicamente nulo, e X_w a STFT de x. Quer-se minimizar

$$D_{X,Y} = \sum_{m=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} |X_w(mS,\omega) - Y_w(mS,\omega)|^2 \, d\omega. \tag{2.4}$$

Pode ser visto que a Equação (2.4) corresponde à soma das potências de todos os quadros de um sinal com espectro $X_w(mS,\omega) - Y_w(mS,\omega)$. Aplica-se, então, o teorema de Parseval, que leva a

$$D_{X,Y} = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} [x_w[mS,l] - y_w[mS,l]]^2. \tag{2.5}$$

Na Equação (2.5), y_w é dado por

$$y_w[mS,l] = \frac{1}{2\pi} \int_{-\pi}^{\pi} Y_w(mS,\omega) e^{j\omega l} d\omega. \tag{2.6}$$

Neste contexto, x_w é a versão janelada do sinal estimado x, o que permite escrever

$$x_w[mS, l] = w[l - mS]x[l]. \tag{2.7}$$

Substituindo (2.7) em (2.5) chega-se a

$$D_{X,Y} = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} [w[l - mS]x[l] - y_w[mS, l]]^2. \tag{2.8}$$

O problema resume-se à escolha dos valores do sinal $x[n]$. De forma a minimizar $D_{X,Y}$, faz-se $\frac{\partial D_{X,Y}}{\partial x[p]} = 0$, $\forall p \in \mathbb{Z}$, ou seja, o gradiente da função é igualado a zero. Como

$$\frac{\partial D_{X,Y}}{\partial x[p]} = 2 \sum_{m=-\infty}^{\infty} w[p - mS][w[p - mS]x[p] - y_w[mS, p]], \tag{2.9}$$

então

$$x[p] = \frac{\sum_{m=-\infty}^{\infty} w[p - mS]y_w[mS, p]}{\sum_{m=-\infty}^{\infty} w^2[p - mS]}, \tag{2.10}$$

que é a equação de atualização utilizada ao longo das iterações. A fim de adequar a notação da Equação (2.10) à explicação do processo iterativo dada em seguida, esta pode ser reescrita como

$$x_i[p] = \frac{\sum_{m=-\infty}^{\infty} w[p - mS]y_{w,i}[mS, p]}{\sum_{m=-\infty}^{\infty} w^2[p - mS]}. \tag{2.11}$$

O processo iterativo é inicializado em $i = 0$ fazendo-se

$$Y_{w,0}(mS, \omega) = |Y_w(mS, \omega)|\,. \tag{2.12}$$

Calcula-se a transformada inversa de $Y_{w,0}$, gerando o sinal $y_{w,0}$. Este é então aplicado à Equação (2.11), concluindo a inicialização com a obtenção do sinal x_0. O processo iterativo que se segue pode ser acompanhado na Figura 2.1.

Na primeira iteração, $i = 1$, é obtida a STFT de x_0, $X_{w,1}$. Cabe ressaltar que a STFT $X_{w,i}$ (calculada na iteração i) é obtida a partir de x_{i-1}. O espectro de fase desta transformada é associado ao espectro de magnitude $|Y_w\,(mS, \omega)|$ através de

$$Y_{w,1}(mS, \omega) = |Y_w(mS, \omega)|\, e^{j\angle X_{w,1}(mS,\omega)}, \tag{2.13}$$

realizando a atualização da MSTFT.

A transformada inversa de $Y_{w,1}$, $y_{w,1}$, é obtida e aplicada à Fórmula (2.11), resultando em uma nova estimativa x_1.

Este processo se repete até que um critério de parada (dos quais alguns tipos são descritos na Seção 2.2.2) seja alcançado. Neste ponto se toma o sinal x_i como resultado.

2.2.2 Implementação

Algumas considerações podem ser feitas para que a estimativa seja calculada com menor custo computacional. Uma consideração feita por Griffin e Lim [3] aborda a utilização de formas para w que possuem a propriedade

$$\sum_{m=-\infty}^{\infty} w^2[n - mS] = 1. \tag{2.14}$$

As formas mostradas derivam das janelas retangular, de Hamming e de Hann, impondo-se pequenas modificações às suas formas tradicionais mostradas nas Equações (2.15), (2.16) e (2.17), respectivamente.

$$w_{\text{retangular}}[n] = \begin{cases} 1, & 0 \leq n < L \\ 0, & \text{caso contrário.} \end{cases} \tag{2.15}$$

$$w_{\text{hamming}}[n] = w_{\text{retangular}}[n] \left[0{,}54 - 0{,}46 \cos\left(\frac{2\pi n}{L-1}\right)\right]. \tag{2.16}$$

$$w_{\text{hann}}[n] = 0{,}5 w_{\text{retangular}}[n] \left[1 - \cos\left(\frac{2\pi n}{L-1}\right)\right]. \tag{2.17}$$

Para a janela retangular, o único ajuste se refere à constante multiplicativa, de modo que podemos utilizar uma janela

$$w_r[n] = \sqrt{\frac{S}{L}} w_{\text{retangular}}[n]. \tag{2.18}$$

No caso das janelas de Hamming e Hann, ambas serão da forma

$$w_h[n] = \frac{2w_r[n]}{\sqrt{4a^2 + 2b^2}} \left[a + b \cos\left(\frac{2\pi n}{L} + \frac{\pi}{L}\right)\right]. \tag{2.19}$$

Fazendo-se $a = 0{,}54$ e $b = -0{,}46$ obtém-se uma janela denominada de Hamming modificada. Ao se fazer $a = 0{,}5$ e $b = -0{,}5$ obtém-se uma janela denominada

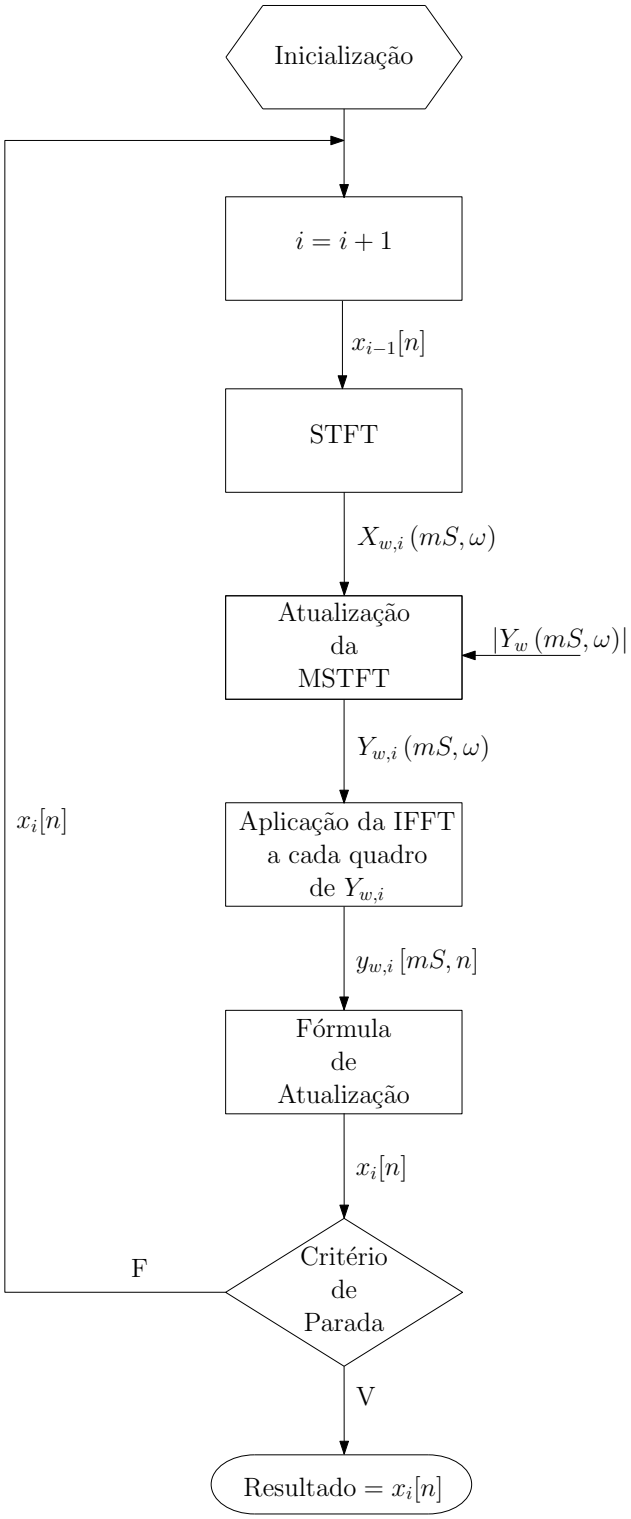

Figura 2.1: Fluxograma do Algoritmo G&L

Hann modificada. Cabe ressaltar que o período da função cosseno contida nesta definição das janelas passa a ser de L amostras, e não as $L-1$ daquelas normalmente encontradas na literatura. Para que as janelas dadas pela Equação (2.19) satisfaçam a Equação (2.14), deve-se ter também $S = L/4$, isto é, 75% de sobreposição entre as janelas, fazendo com que o denominador da Equação (2.10) possa ser substituído por 1.

Um outro ponto a ser examinado é a eficácia do método. De forma a ser possível quantificá-la, pode-se avaliar o quanto o espectro de magnitude obtido através da STFT de x se aproxima de $|Y_w|$. Griffin e Lim [3] propuseram uma medida de distância dada por

$$D_{M_{X,Y}} = \sum_{m=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} [|X_w(mS,\omega)| - |Y_w(mS,\omega)|]^2 \, d\omega. \tag{2.20}$$

Notar que a Equação (2.4) consiste na distância entre uma STFT e uma MSTFT, enquanto, através da Equação (2.20), é medida a distância entre espectros de magnitude, o que é denotado pelo índice M. É demonstrável que a cada iteração do método, a medida da Equação (2.20) decresce [3]. A melhora obtida, no entanto, torna-se menos significativa com o aumento das iterações, chegando a um ponto a partir do qual a melhora não compensa o custo computacional. Essa constatação sugere um critério de parada, proposto neste trabalho, alternativo ao estabelecimento de um número máximo de iterações, como proposto por Griffin e Lim. Com o novo critério, descrito a seguir, é possível utilizar o algoritmo de forma mais eficiente.

De posse de uma nova medida de distância ao fim de cada iteração, calcula-se a variação entre a última medida e a anterior a fim de concluir até que ponto o aumento do número de iterações influencia, com a significância desejada, o sinal estimado. No caso da implementação adotada neste trabalho, o critério consistiu na escolha de um limiar de variação percentual, $t(\%)$. Seja D_{M_i} a distância correspondente à iteração i. O critério consiste em avaliar se

$$\frac{D_{M_{i-1}} - D_{M_i}}{D_{M_{i-1}}} < t(\%). \tag{2.21}$$

Se a Inequação (2.21) é satisfeita, a execução do método é interrompida e x_i é tomado como resultado final.

Como as operações são realizadas em computadores, é necessária a utilização de uma representação discreta na frequência. Utiliza-se, então, a Transformada

Discreta de Fourier (DFT, do inglês *Discrete Fourier Transform*), calculada através do algoritmo rápido FFT (do inglês *Fast Fourier Transform*). Para o cálculo da transformada inversa, calcula-se a Transformada Discreta de Fourier Inversa (IDFT, do inglês *Inverse Discrete Fourier*), calculada usando o algoritmo rápido IFFT (do inglês *Inverse Discrete Fourier Transform*). Como mostrado na Figura 2.1, estes algoritmos são usados tanto no cálculo da STFT como no passo *Aplicação da IFFT a cada quadro de* Y_{w_i}. Estes algoritmos rápidos foram utilizados para os cálculos das representações discretas na frequência também para os demais algoritmos de reconstrução de fase.

2.3 RTISI (*Real-Time Iterative Spectrogram Inversion*)

2.3.1 Desenvolvimento Teórico

No algoritmo G&L, a estimação de um sinal temporal depende da disponibilidade da MSTFT Y_w para todos os quadros do sinal a ser estimado, uma vez que m varia desde $-\infty$ a ∞ na Equação (2.10). Com base nesta condição, pode-se perceber a impossibilidade de se utilizar o algoritmo G&L para aplicações em tempo real. Nos algoritmos com este propósito, o índice m pode variar até no máximo o valor Q, que se refere ao quadro corrente.

Para contornar esta limitação, Beauregard *et al.* [4] propuseram um algoritmo no qual o sinal temporal é construído quadro a quadro. Este algoritmo, denominado Inversão Iterativa de Espectrograma em Tempo Real (RTISI), aborda a reconstrução 'em série', calculando os quadros do sinal sequencialmente, em constraste com a reconstrução paralela de todos os quadros realizada no G&L.

É importante ressaltar que a definição de tempo real utilizada refere-se à possibilidade de estimação da fase do quadro Q. A estimação deve ser feita apenas com informações já dadas até o instante em que seu espectro de magnitude está disponível.

Seja $y_w(mS, n)$ uma função tal que se tenham valores dos quadros já reconstruídos para $m < Q$ e $y_w(QS, n) = 0 \ \forall n$. De forma a construir o quadro Q do

sinal, é efetuada a minimização de um erro quadrático médio, tal como no algoritmo G&L. Neste caso, porém, em vez de se minimizar a Equação (2.4), é realizada a minimização de

$$D_{X,Y} = \sum_{m=-\infty}^{Q} \sum_{l=-\infty}^{\infty} [w[l - mS]x[l] - y_w[mS, l]]^2 \tag{2.22}$$

em relação aos valores de $x[n]$. De forma análoga à minimização da Equação (2.4), chega-se a

$$x[n] = \frac{\sum_{m=-\infty}^{Q} w[n - mS]y_w[mS, n]}{\sum_{m=-\infty}^{Q} w^2[n - mS]}, \tag{2.23}$$

com $QS \leq n \leq (L-1) + QS$, que será a fórmula de atualização neste caso.

A Equação (2.23) pode ser reescrita de modo a enfatizar o termo que sofrerá mudanças durante o processo iterativo, da seguinte forma:

$$x[n] = \frac{w[n - QS]y_w[QS, n]}{\sum_{m=-\infty}^{Q} w^2[n - mS]} + \frac{\sum_{m=-\infty}^{Q-1} w[n - mS]y_w[mS, n]}{\sum_{m=-\infty}^{Q} w^2[n - mS]}. \tag{2.24}$$

Define-se aqui $x_{\text{parcial}}[mS, n]$ como a função que armazena os quadros já estimados do sinal, de tal forma que $x_{\text{parcial}}[mS, n] = y_w[mS, n]$ para $m < Q$.

Seja, agora

$$y_{\text{RTISI}}[q, n] = \sum_{m=-\infty}^{q} w[n - mS]x_{\text{parcial}}[mS, n]. \tag{2.25}$$

Observando-se as Equações (2.24) e (2.25), pode-se concluir que

$$x[n] = \frac{w[n - QS]y_w[QS, n] + y_{\text{RTISI}}[Q-1, n]}{\sum_{m=-\infty}^{Q} w^2[n - mS]}. \tag{2.26}$$

Por fim, ao aplicar a Equação (2.24), são obtidas estimativas de amostras do sinal temporal resultante no intervalo de tempo do quadro corrente. Multiplica-se, então, este sinal $x[n]$ pela janela w, obtendo-se a estimativa $x_w[QS, n]$ do quadro corrente Q através de

$$x_w[QS, n] = w[n - QS]x[n]. \tag{2.27}$$

O fluxograma da Figura 2.2 representa o processo iterativo do algoritmo para o quadro Q. Na inicialização do processo faz-se $i = 0$ e $Y_{w,0}(QS, \omega) = |Y_w(QS, \omega)|$. A partir de $Y_{w,0}(QS, \omega)$ obtém-se $y_{w,0}$, que ao ser aplicado à Fórmula (2.26) leva à obtenção de um sinal $x_0[n]$. Este sinal é então multiplicado pela janela, como na Equação (2.27), resultando na estimativa inicial do quadro Q, $x_{w,0}[QS, n]$, e concluindo a inicialização do processo.

Assim como no algoritmo G&L, a transformada $X_{w,i}(QS, \omega)$ (calculada na iteração i) é obtida a partir de $x_{w,i-1}[QS, n]$. O processo é repetido até que se atinja um certo critério de parada (discutido na Seção 2.3.2). Após a estimativa do quadro Q, estima-se o quadro $Q + 1$ do mesmo modo.

O resultado do algoritmo, $x_{\text{final}}[n]$, é obtido pela sobreposição dos quadros estimados:

$$x_{\text{final}}[n] = \frac{\sum_{m=-\infty}^{\infty} x_{\text{parcial}}[mS, n]}{\sum_{m=-\infty}^{\infty} w[n - mS]}. \tag{2.28}$$

Este processo de sobreposição pode ser realizado de modo que se tenham S amostras disponíveis a cada quadro estimado, ou pode ser realizado ao fim da estimativa de todos os quadros do sinal. Neste trabalho foi adotada a sobreposição de todos os quadros ao fim da estimativa.

Por utilizar apenas informações de quadros passados e do quadro corrente para a estimativa do quadro corrente, espera-se que, para um número elevado de iterações, este algoritmo produza resultados com menor qualidade do que os resultados do algoritmo G&L, que utiliza informações de todos os quadros.

2.3.2 Implementação

A implementação do algoritmo seguiu a Fórmula (2.23). Beauregard *et al.* [4] propõem que as janelas propostas por Griffin e Lim, mostradas na Seção 2.2.2, sejam utilizadas. Diante dos cálculos efetuados neste algoritmo, no entanto, não se vê motivo para isto, uma vez que o cumprimento da Condição (2.14) não leva à redução do número de operações envolvidas no cálculo da Equação (2.23). Notar que a Condição (2.14) tem os índices do somatório diferentes daqueles da Equação (2.23), não permitindo que o denominador desta equação seja substituído pela unidade, como

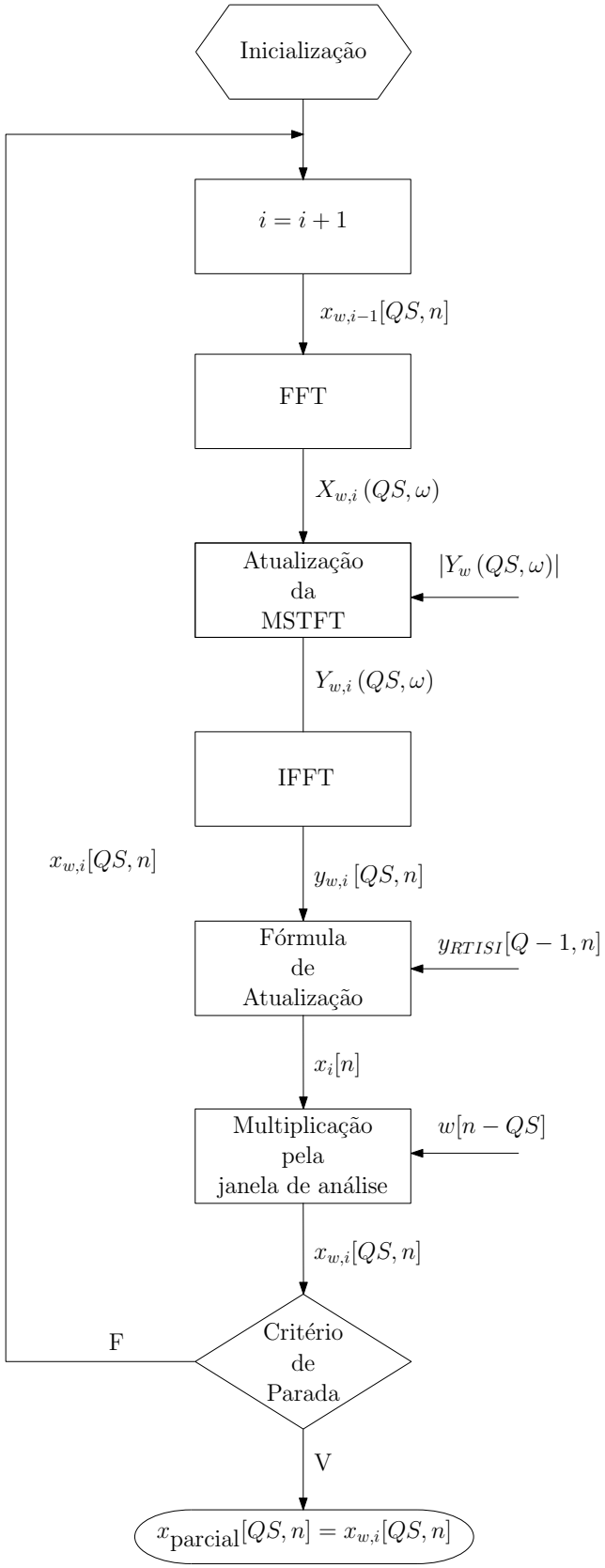

Figura 2.2: Fluxograma da estimativa do quadro Q através do Algoritmo RTISI

no algoritmo G&L. Contudo, para realizar comparações justas entre este algoritmo e o G&L, as mesmas janelas foram utilizadas.

Beauregard *et al.* propõem também o estabelecimento de um critério de parada dado por um número máximo de iterações a ser realizado sobre cada quadro, sendo o mesmo número para todos os quadros. Porém, em quadros nos quais haja menos informação no sinal como, por exemplo, em trechos de silêncio ou trechos sem ataques de instrumentos, esse critério pode levar a um excesso de operações.

Para resolver esta questão, neste trabalho é proposta a utilização de um critério de parada análogo ao da Inequação (2.21). Todavia, em se tratando apenas do quadro Q, o cálculo da medida é efetuado por

$$D_{M_{X,Y}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[|X_w(QS, \omega)| - |Y_w(QS, \omega)| \right]^2 d\omega. \tag{2.29}$$

Notar que, neste caso, apenas a distância do quadro corrente Q é considerada, diferentemente do que ocorre na Equação (2.20). Ao se adotar este critério, o número de operações pode ser diferente para cada quadro, o que permite levar em conta características de cada trecho do sinal.

2.4 RTISI-LA (*Real-Time Iterative Spectrogram Inversion with Look Ahead*)

2.4.1 Desenvolvimento Teórico

O desenvolvimento do algoritmo RTISI, apresentado na Seção 2.3, levou à obtenção de um algoritmo que atende à propriedade de reconstrução em tempo real. Apesar disto, sabe-se que, devido à superposição entre quadros consecutivos, alguns quadros posteriores possuem informação sobre o quadro corrente Q. Isto é ilustrado na Figura 2.3 para $L = 1024$ e $S = 256$. Nesse caso, como há 75% de sobreposição, 3 quadros posteriores compartilham informação com o quadro Q. Ao desprezar essa contribuição, o RTISI adquire uma limitação para suas estimativas.

Com o intuito de contornar essa limitação, Beauregard *et al.* [8] propuseram a flexibilização da característica de tempo real. Isto foi feito através da utilização de K quadros posteriores do sinal na reconstrução do corrente Q, fato denotado

pela sigla LA (do inglês *Look Ahead*) no nome deste algoritmo, que significa olhar à frente.

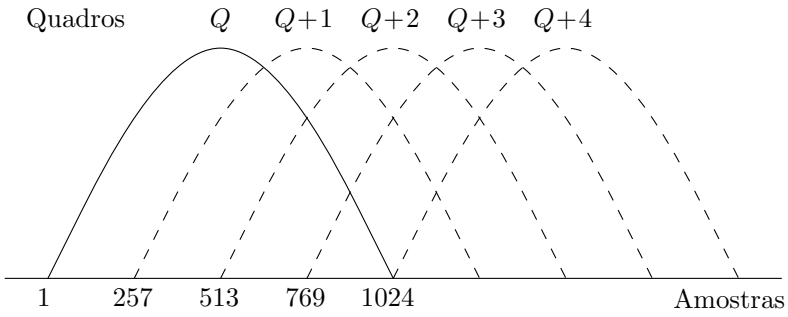

Figura 2.3: Superposição de quadros posteriores no quadro corrente Q

Essa flexibilização se dá pela inserção destes K quadros na Fórmula (2.23), resultando em uma equação na qual $x[n]$ é dado por

$$x[n] = \frac{\sum_{m=-\infty}^{Q+K} w[n-mS]y_w[mS,n]}{\sum_{m=-\infty}^{Q+K} w^2[n-mS]}, \tag{2.30}$$

que passa a ser a nova fórmula de atualização.

Deve-se notar que, embora na Fórmula (2.30) o cálculo deva ser efetuado para $QS \leq n \leq (L-1)+(Q+K)S$, apenas o quadro Q é guardado como parte do resultado final ao fim das iterações para esse quadro.

Como efeito da mudança no cálculo, o método se mostra mais robusto, levando em conta informações de quantos quadros posteriores a Q se desejar. Embora os cálculos de cada quadro envolvam outros quadros além dele próprio, este algoritmo mantém a principal característica do RTISI: realizar a reconstrução de fase quadro a quadro de forma sequencial.

Utilizando o mesmo desenvolvimento e definições do algoritmo RTISI original, a fórmula de atualização passa a ser escrita, de forma análoga à Equação (2.26), como

$$x[n] = \frac{\sum_{m=Q}^{Q+K} w[n-mS]y_w[mS,n] + y_{\text{RTISI}}[Q-1,n]}{\sum_{m=-\infty}^{Q+K} w^2[n-mS]}. \tag{2.31}$$

A inicialização do processo é feita com $Y_{w,0}(mS,\omega) = |Y_w(mS,\omega)|$, com m variando de Q a $Q+K$ neste caso. Obtém-se a transformada inversa de cada quadro

de $Y_{w,0}(mS, \omega)$, $y_{w,0}$. Em seguida, o sinal $x_0[n]$ é avaliado, a partir da Equação (2.30), para $1 + (Q-1)S \leq n \leq L + (Q+K-1)S$, sendo obtidas $L + KS$ amostras. Estas amostras são janeladas, gerando estimativas iniciais para os quadros, as quais são armazenadas em $x_{w,0}$, terminando a inicialização. O restante do processo é então executado como ilustrado na Figura 2.4, onde $1+(Q-1)S \leq n \leq L + (Q+K-1)S$ e $Q \leq m \leq Q + K$.

Tal como feito na Seção 2.3, a sobreposição dos resultados é realizada apenas após a estimação de todos os quadros do sinal.

2.4.2 Implementação

Ao fim da estimação do quadro Q, há estimativas parciais feitas para seus K quadros seguintes, de $Q+1$ a $Q+K$. Estas estimativas podem ser tratadas de dois modos. No primeiro, armazena-se as K estimativas parciais de forma a inicializar seus valores na estimação do quadro $Q+1$. No segundo, inicializa-se com fase identicamente nula, ou seja, $Y_{w,0}(mS, \omega) = |Y_w(mS, \omega)|$ para $Q+1 \leq m \leq Q+K$. Em qualquer dos casos, como não há nenhuma estimativa para a fase do quadro $Q+K+1$, ele é inicializado com $Y_{w,0}((Q+K+1)S, \omega) = |Y_w((Q+K+1)S, \omega)|$.

Na implementação do algoritmo, o primeiro modo foi adotado, pois tende a reduzir a distância dada pela Equação (2.29) com um número menor de iterações. Isto ocorre porque as informações dos quadros anteriores já terão sido utilizadas na estimativa inicial dos quadros de $Q+1$ a $Q+K$.

2.5 MISI

2.5.1 Desenvolvimento Teórico

Os algoritmos G&L, RTISI e RTISI-LA buscam reconstruir a fase de um sinal com base em informações referentes apenas ao seu espectrograma. No entanto, em casos em que o espectrograma é proveniente de processos de separação de fontes, o sinal da mistura pode ser de grande valia no processo de reconstrução. A estrutura do algoritmo MISI [5], que utiliza esse sinal para realizar as estimativas, será apresentada nesta seção.

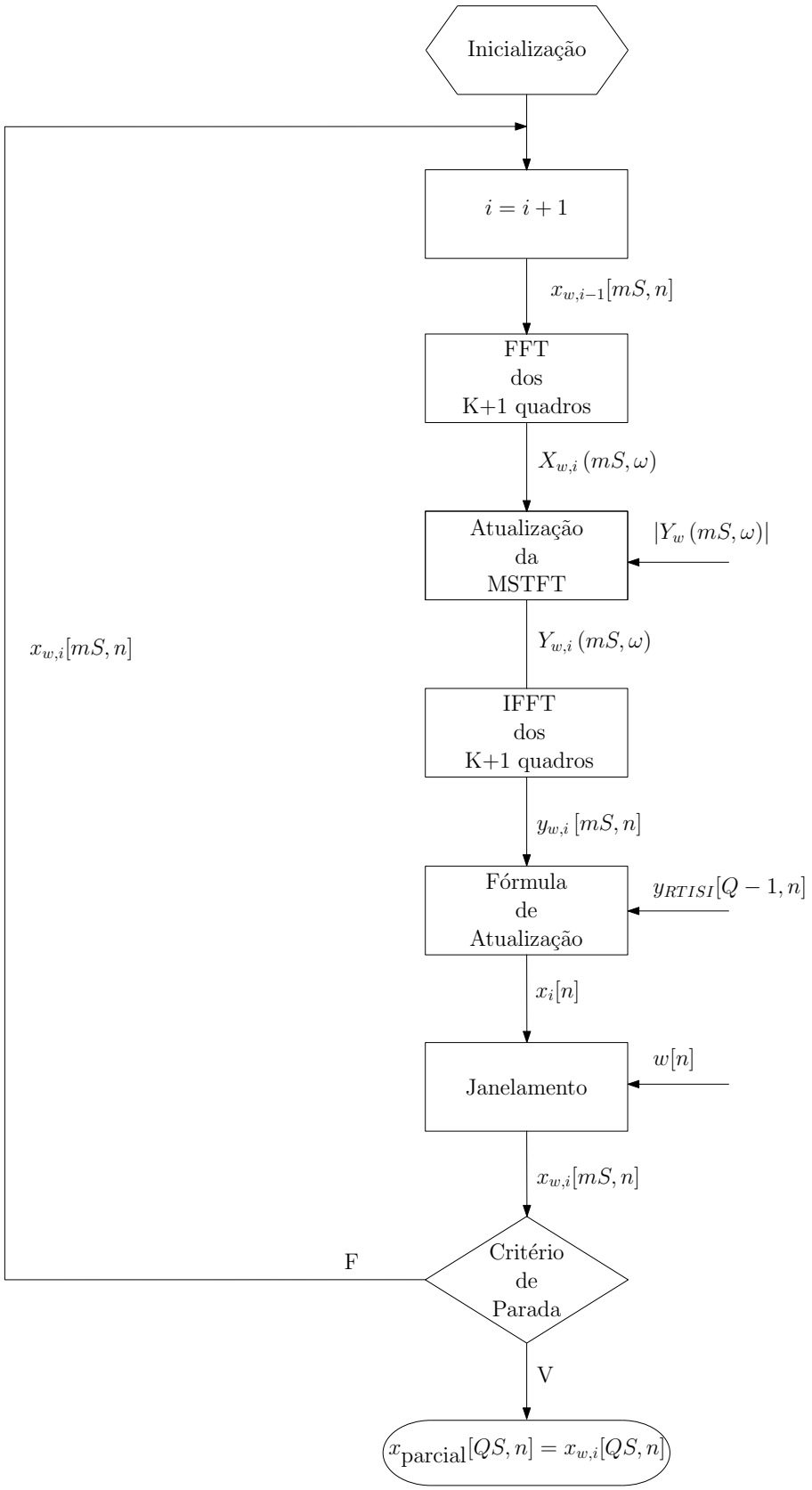

Figura 2.4: Fluxograma da estimativa do quadro Q através do Algoritmo RTISI-LA

Como um exemplo mais simples da utilidade do sinal de mistura na obtenção das fases individuais das fontes, pode-se pensar em dois instrumentos que sejam tocados em trechos alternados. Tendo disponível o sinal da mistura e obtendo os espectrogramas das fontes, o espectro de fase de cada instrumento será de fácil obtenção, uma vez que cada instrumento deverá aparecer sozinho ao longo de todos os quadros da mistura. Embora este exemplo seja de um caso bastante particular, é comum que as fontes em uma mistura não estejam todas ativas ao longo de todo o sinal de mistura, o que propicia uma melhor estimação das fases.

Com base nesta ideia, a utilização da fase da mistura na inicialização dos métodos anteriores já fornece informação sobre a fase das fontes.

Gunawan e Sen [5], supondo que a mistura das fontes é puramente aditiva, propuseram que a cada iteração os sinais temporais estimados para as fontes fossem somados, e esta soma subtraída do sinal da mistura dada, gerando um sinal de erro $e_i[n]$. A partir desta ideia foi inserida uma pequena modificação em relação ao algoritmo G&L. Sendo $x_{\text{fonteF},i-1}$ o sinal temporal estimado para uma fonte F na iteração $i-1$, em vez de utilizar a fase deste sinal para o passo *atualização da MSTFT*, é proposto que se utilize a fase correspondente à STFT de

$$r_{\text{fonteF},i}[n] = x_{\text{fonteF},i-1}[n] + \frac{e_i[n]}{P}, \tag{2.32}$$

onde P é o número de fontes.

Muitos dos passos utilizados no algoritmo G&L são utilizados no MISI. Assim, pode-se definir um bloco de processos, chamado *Núcleo G&L*, que engloba esses passos. O bloco é mostrado na Figura 2.5.

Definido o bloco *Núcleo G&L*, a estrutura do algoritmo MISI pode ser representada como na Figura 2.6. Na inicialização do algoritmo, o espectro de magnitude de cada fonte é associado à fase da mistura, gerando novas MSTFTs para cada fonte. A partir destas MSTFTs são obtidos os sinais $x_{\text{fonteF},0}$, terminando a inicialização. Na Figura 2.6, o algoritmo é ilustrado para o caso de duas fontes, sendo o caso para P fontes análogo.

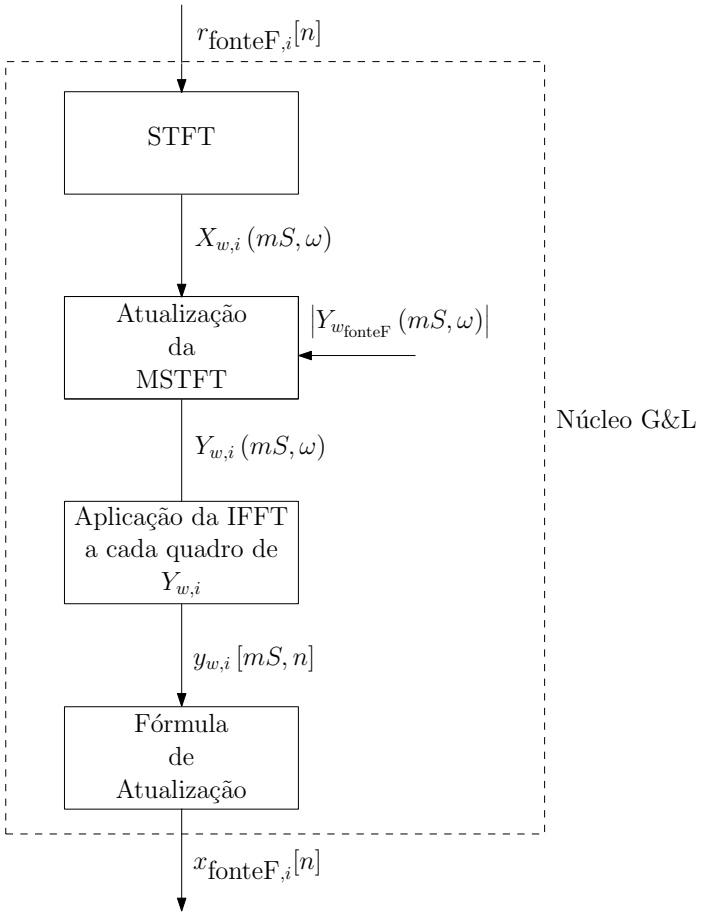

Figura 2.5: Definição do bloco *Núcleo G&L*

2.5.2 Implementação

Por utilizar um desenvolvimento análogo ao do algoritmo G&L, são também aplicáveis ao algoritmo MISI as janelas modificadas e o critério de parada pela variação da distância descritas na Seção 2.2.2.

2.6 Visão Geral dos Métodos

Feita a discussão da base teórica dos quatro métodos de reconstrução retratados neste trabalho, aspectos gerais dos métodos podem ser ressaltados de forma comparativa.

Entre os métodos G&L e RTISI, há uma ponderação entre desempenho e custo computacional. Pode-se pensar que o método G&L tenha a capacidade de gerar melhores estimativas que o RTISI, tendo, no entanto, um maior custo compu-

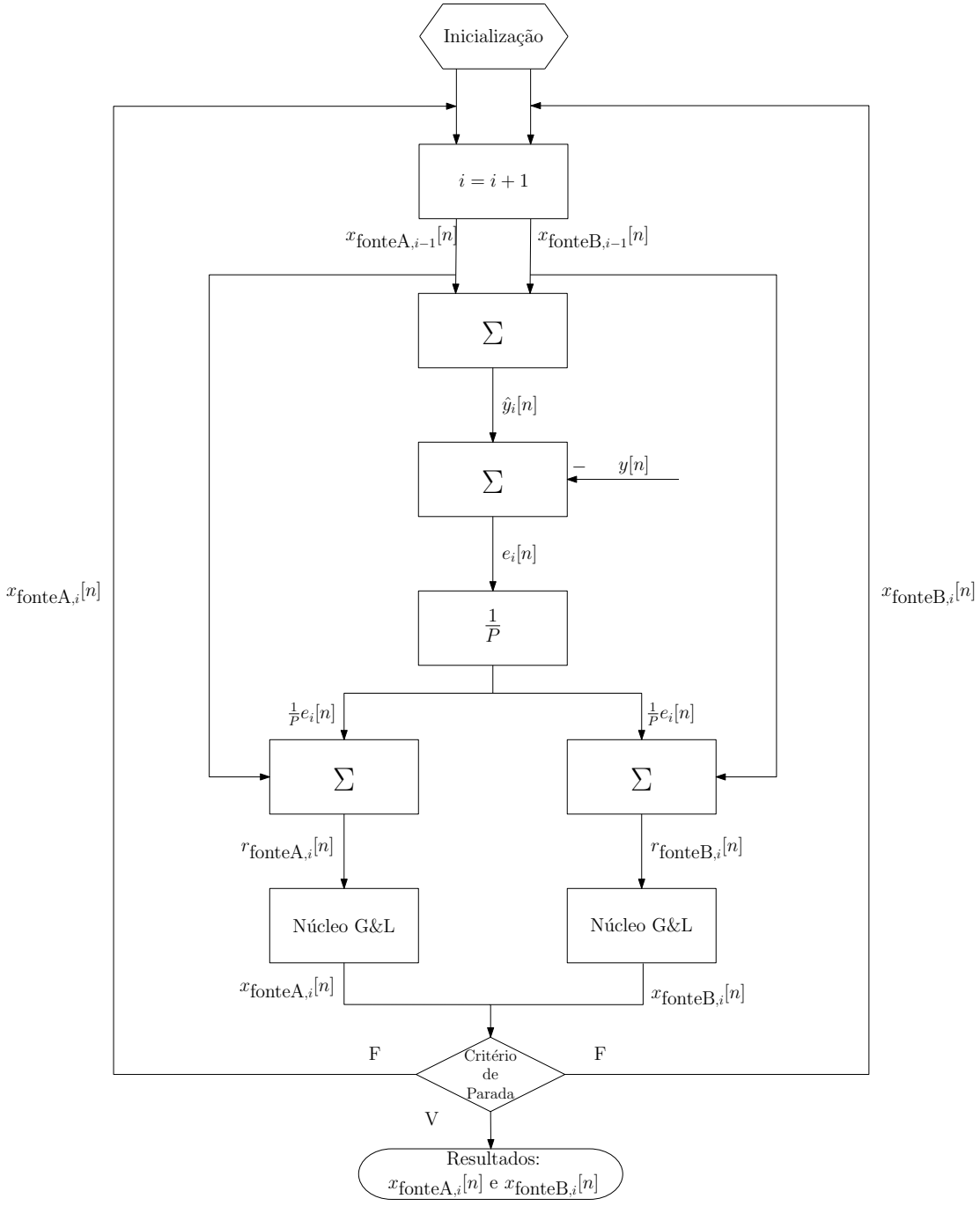

Figura 2.6: Fluxograma do Algoritmo MISI para 2 fontes ($P = 2$)

tacional.

Quanto ao método RTISI-LA, a escolha do parâmetro K influencia o desempenho da estimação e estabelece o número de operações realizadas a cada iteração. Assim, escolhido $K > 0$, o RTISI-LA não recai na limitação do método RTISI: não utilizar a informação de quadros posteriores na estimação do quadro corrente. Por outro lado, a utilização de quadros posteriores aumenta o número de operações realizadas a cada iteração com relação ao método RTISI. Em aplicações de tempo real, este método impõe a necessidade de uma latência. Porém, não é necessária a disponibilidade de todo o sinal, como no algoritmo G&L, para que a estimação seja realizada.

O método MISI tem por base o método G&L. No entanto, este método utiliza um sinal de mistura para auxiliar na reconstrução dos sinais de suas fontes. É proposta a inicialização da fase dos sinais das fontes com a fase do sinal da mistura e a utilização deste sinal para o cálculo de um sinal de erro. Este sinal de erro é utilizado para aprimorar o processo de estimação.

Capítulo 3

Resultados Comparativos

3.1 Metodologia de Comparação

No Capítulo 2 foi apresentada uma medida para o desempenho dos métodos, repetida aqui por conveniência:

$$D_{M_{X,Y}} = \sum_{m=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[|X_w(mS,\omega)| - |Y_w(mS,\omega)| \right]^2 d\omega, \tag{3.1}$$

onde $X_w(mS,\omega)$ é a STFT do sinal estimado e $Y_w(mS,\omega)$ é a MSTFT obtida a partir do espectrograma dado.

Outra medida foi proposta por Gunawan e Sen [5], e é dada por

$$\mathrm{SNR} = 10 \log_{10} \left(\frac{\sum_{n=-\infty}^{\infty} (x[n])^2}{\sum_{n=-\infty}^{\infty} (x[n] - \hat{x}[n])^2} \right), \tag{3.2}$$

sendo $x[n]$ o sinal original da fonte e $\hat{x}[n]$ o sinal estimado.

A primeira depende da energia do sinal, não sendo uma boa medida de qualidade, uma vez que varia com sua normalização.

A segunda possui dois pontos que a desqualificam como medida prática de comparação. O primeiro é a necessidade de se conhecer o sinal original, algo incoerente com as situações práticas de aplicação dos métodos. O segundo diz respeito ao cálculo do denominador: estando disponível apenas o espectro de magnitude, o problema de encontrar um sinal $\hat{x}[n]$ que possua esse espectro como magnitude de

sua MSTFT não possui solução única. Pode-se encontrar um sinal $\hat{x}[n]$ perceptivamente similar a $x[n]$ sem que estes possuam formas de onda similares. Neste caso, por causa da semelhança perceptiva entre os sinais, seria interessante aplicar uma medida que indicasse um bom resultado. No entanto, a medida da SNR (do inglês, *Signal-to-Noise Ratio*), que considera ser ruído a diferença entre o sinal temporal estimado $\hat{x}[n]$ e o sinal original $x[n]$, será baixa.

Assim sendo, nenhuma dessas duas medidas é apropriada para comparação dos resultados dos diferentes métodos. É desejável, então, um parâmetro que dependa apenas das informações disponíveis no início dos processos de reconstrução de fase e das estimativas do sinal temporal.

No fim dos processos, há dois espectros de magnitude. O espectro obtido da MSTFT através do espectrograma dado, como descrito na Seção 2.2, denotado por $|Y_w(mS,\omega)|$, e o espectro $|X_w(mS,\omega)|$, obtido da STFT da estimativa do sinal. Estes são os únicos parâmetros necessários no cálculo da SER (do inglês, *Signal-to-Error Ratio*), medida citada por Beauregard *et al.* [4]:

$$\mathrm{SER} = -10\log_{10}\left(\frac{\displaystyle\sum_{m=-\infty}^{\infty}\int_{-\pi}^{\pi}\left[|X_w(mS,\omega)| - |Y_w(mS,\omega)|\right]^2 d\omega}{\displaystyle\sum_{m=-\infty}^{\infty}\int_{-\pi}^{\pi}|Y_w(mS,\omega)|^2\, d\omega}\right). \tag{3.3}$$

Neste capítulo, os desempenhos dos métodos serão avaliados segundo os resultados desta medida.

Observando a Equação (2.20) [3], é possível ver que a SER pode ser reescrita como

$$\mathrm{SER} = -10\log_{10}\left(\frac{D_{M_{X,Y}}}{\dfrac{1}{2\pi}\displaystyle\sum_{m=-\infty}^{\infty}\int_{-\pi}^{\pi}|Y_w(mS,\omega)|^2\, d\omega}\right). \tag{3.4}$$

Deve-se notar que, por utilizar um $D_{M_{X,Y}}$ normalizado, esta medida independe da potência do espectro de magnitude dado.

Como $D_{M_{X,Y}}$ é uma medida de distância entre espectros de magnitude, quanto mais próxima for a estimativa do sinal representado por $|Y_w(mS,\omega)|$, maior deve ser o valor da SER, e melhor a qualidade do sinal resultante.

3.2 Testes

Para que os testes possam ser abrangentes é necessário que sejam feitos utilizando sinais com características diversas. Neste trabalho, foram escolhidos 13 sinais para testes com os algoritmos G&L, RTISI e RTISI-LA.

Esses sinais foram divididos em Sinais Tonais, Sinais Percussivos e Sinais Tonais e Percussivos. Os sinais possuem 16 bits e frequência de amostragem de 44100 Hz com durações que variam entre 5 e 15 segundos, exceto por um da terceira família, que tem frequência de amostragem de 22050 Hz e duração de 24 segundos.

Foram realizadas tentativas de reconstrução de fase usando diferentes números de iterações. Os algoritmos foram inicializados com fase nula em todos os casos. Buscou-se também verificar a dependência da qualidade dos resultados com o número de pontos da DFT.

Para o método RTISI-LA, foram testados diferentes números de quadros de *Look Ahead* tanto para os Sinais Tonais quanto para os Percussivos. Por estes dois tipos serem bastante representativos quanto à variedade de características de sinais, esse teste não foi realizado com os demais sinais.

A fim de traçar uma análise comparativa com o método MISI, além dos sinais citados anteriormente, foram utilizados Sinais de Fontes Pré Mixagem, disponíveis no *site* da SiSEC 2010 (*http://sisec.wiki.irisa.fr/tiki-index.php*). O método MISI foi comparado com os demais e o seu desempenho testado para misturas com diferentes números de fontes.

Em todos os testes foi utilizada a janela de Hamming modificada, com duração de 23,2 ms e 75% de sobreposição, a menos que algo diferente seja especificado. Deve-se ressaltar que esta duração de janela corresponde a 1024 pontos para a amostragem de 44100 Hz.

Outro aspecto a ser considerado é o cálculo da SER. Não é possível realizar o cálculo da integral dada no numerador da Equação (3.3), uma vez que a STFT só pode ter um número finito de pontos e a integral não pode ser calculada indiretamente. Portanto, é calculado um valor aproximado da medida: ao se obter um sinal x resultante de uma reconstrução, a sua STFT discreta $X[mS, k]$ é calculada. O número de pontos N empregados na obtenção de $X[mS, k]$ é o mesmo utilizado

na obtenção da MSTFT discreta $Y_w[mS, k]$. Ambas são então aplicadas à Fórmula

$$\mathrm{SER} \approx -10 \log_{10} \left(\frac{\sum_{m=-\infty}^{\infty} \sum_{k=0}^{N-1} \left[|X_w[mS, k]| - |Y_w[mS, k]| \right]^2}{\sum_{m=-\infty}^{\infty} \sum_{k=0}^{N-1} |Y_w[mS, k]|^2} \right). \tag{3.5}$$

Em situações práticas, o número de pontos empregados no cálculo da SER é o mesmo do espectrograma dado. Neste trabalho, esta medida foi calculada para $N = 2^{14}$ para todos os testes, utilizando janela de Hamming modificada de 23,2 ms com sobreposição de 75%. A escolha desse N decorre de este número ser potência de 2 (o que favorece o cálculo da DFT pelo algoritmo FFT) e maior que 10 vezes o número de pontos da janela. Assim, com N muito maior que o número de pontos da janela, pode-se considerar que a integral do numerador será bem aproximada.

Pode-se observar que a realização do cálculo da SER para este N elevado depende da disponibilidade dos sinais originais, um dos argumentos dados para a não utilização da medida de SNR. Porém, na realização dos testes, optou-se por isso para possibilitar a comparação de todos os sinais com base em uma mesma medida.

Além disso, avaliou-se o efeito sobre a aproximação da SER causado por se utilizar N igual ao próprio número de pontos das DFTs que compõem o espectrograma em lugar de $N = 2^{14}$.

Quanto aos valores obtidos para a SER, verificaram-se variações entre 0,5 e 1 unidade da medida por conta de erros numéricos. Foi observado que os valores da SER podem apresentar resultados diferentes, dependendo da precedência de operações escolhida nos algoritmos de estimação de fase. Estas diferenças são devidas à realização dos cálculos com números com precisão limitada. Por exemplo, para obter menor custo computacional na implementação de um algoritmo, pode ser necessário que, para a adição de frações de mesmo denominador, se realize primeiramente as operações de divisão. Entretanto, para minimizar a possibilidade de erros, primeiro seus numeradores deveriam ser somados para depois se realizar a divisão. Devido à observação desses erros, nas análises que se seguem, resultados que apresentarem variações de até 1 unidade da medida serão considerados como tendo o mesmo desempenho.

3.2.1 Sinais Tonais

3.2.1.1 Resultados

Primeiramente, foram observados os resultados para 5 sinais, classificados como Sinais Tonais. A descrição dos sinais é apresentada na Tabela 3.1.

Tabela 3.1: Descrição dos Sinais Tonais

Código	Descrição
T1	Trecho de um *Adagio* para órgão de Johann Sebastian Bach transcrito para violoncelo e piano. Eles soam simultaneamente; o violoncelo usa leve vibrato, e no piano são tocados acordes.
T2	Trecho da peça *Litanies* para órgão de Jehan Alain. O instrumento soa apenas um acorde em ambiente com forte reverberação. Sinal bastante estacionário.
T3	Trecho de um Concerto para cravo e cordas de Johann Sebastian Bach. O solista e a orquestra soam simultaneamente. O andamento é *allegro*.
T4	Trecho da peça *Moto Perpetuo* de Niccolò Paganini transcrita para violoncelo e piano. O violoncelo toca uma rápida sucessão de notas, e no piano são tocados acordes em *staccato*. O andamento é *allegro*.
T5	Trecho da Sonata para piano no. 5 de Alexander Scriabin. O andamento é *presto*, sofrendo leve aumento progressivo em conjunto com a dinâmica; em seguida, o andamento cai e a dinâmica se mantém em *forte*.

De forma a validar as implementações dos métodos de reconstrução, foram realizados testes análogos aos realizados por Beauregard *et al.* [8]. O critério de parada foi o número máximo de iterações e o número de pontos da MSTFT foi coincidente com o número L de pontos da janela ($L = 1024$ para os Sinais Tonais). Buscou-se comparar como os diferentes métodos se comportam com a variação do

número de iterações. Os resultados estão representados na Figura 3.1.

Pode-se observar que o método RTISI tem um desempenho superior ao do método G&L para poucas iterações, sendo superado com o aumento deste número. Este fato, também observado por Beauregard *et al.*, ocorreu na maioria dos casos testados.

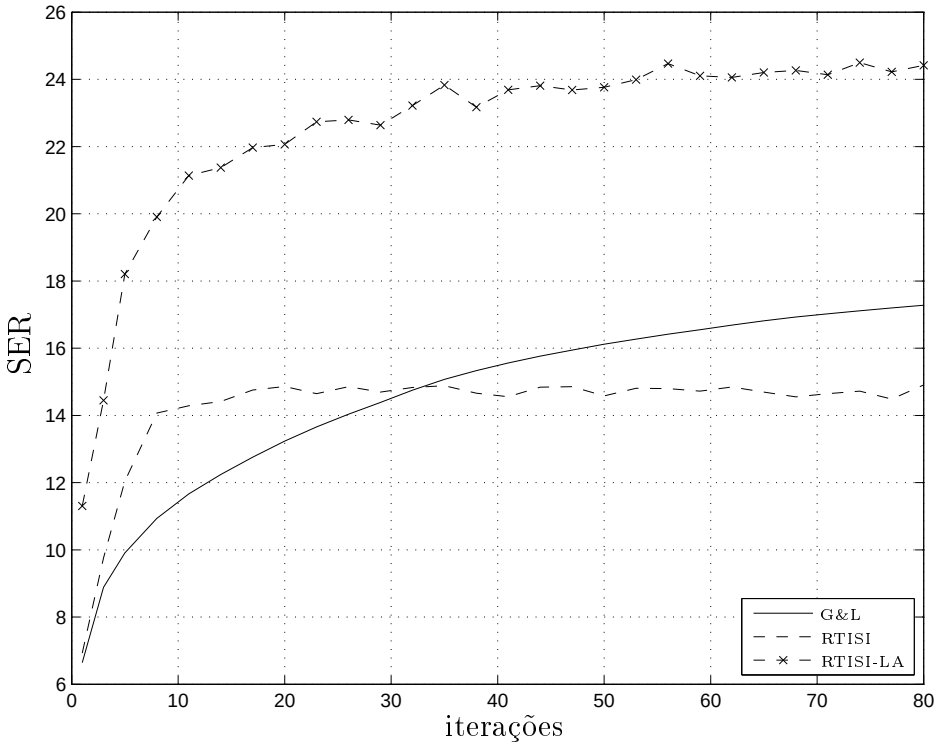

Figura 3.1: Resultados para o Sinal T3 com DFT de 1024 pontos

O desempenho do RTISI para poucas iterações, em geral superior, sugere que com este método se obtém uma melhor estimativa inicial de fase para o quadro corrente. Essa melhor estimativa ocorre porque o termo y_{RTISI}, definido na Seção 2.3, já contém estimativas prontas dos quadros anteriores. No método G&L, diferentemente, todos os quadros são estimados paralelamente.

Ao utilizar apenas informações de quadros anteriores, o método RTISI atingiu rapidamente uma assíntota para a qualidade de reconstrução, fato que pode ser observado tanto na Figura 3.2, como na Figura 3.1. Após atingir essa assíntota, o desempenho do algoritmo RTISI acabou sendo superado pelo algoritmo G&L em todos os casos.

A exceção para o melhor desempenho do RTISI sobre o G&L para poucas iterações ocorreu com o Sinal T2. Neste caso, o método G&L teve um desempe-

nho superior para todos os números de iterações testados, como pode ser visto na Figura 3.2. Para este sinal, o método RTISI atinge a assíntota do seu desempenho com poucas iterações. Isto pode ocorrer devido à forte estacionariedade de T2, que favorece uma rápida convergência dos resultados da Equação (2.26). O método G&L consegue, para cada quadro, extrair a informação de todos os quadros vizinhos. Com a estacionariedade também favorecendo à extração mais rápida da informação e com mais informação disponível, método G&L consegue, então, superar a vantagem da melhor estimativa inicial de fase do método RTISI.

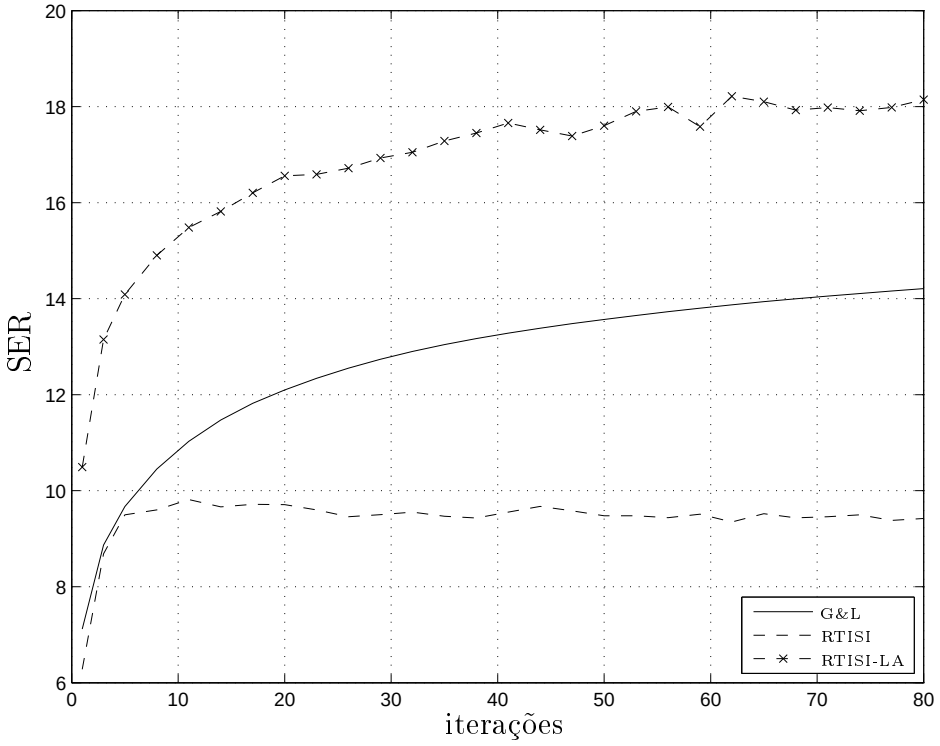

Figura 3.2: Resultados para o Sinal T2 com DFT de 1024 pontos

Contando com a melhor estimativa inicial, tal como no método RTISI, e utilizando K quadros posteriores ao quadro corrente, o método RTISI-LA apresentou, em todos os testes realizados com Sinais Tonais, desempenho bastante superior aos outros dois métodos. Nestes testes foram utilizados apenas 3 quadros de *Look Ahead*.

Vista a diferença de desempenho entre os métodos RTISI e RTISI-LA, investigou-se quantos quadros de *Look Ahead* seriam necessários para que o RTISI-LA superasse o desempenho do G&L. Além disso, buscou-se averiguar a variação de desempenho em função do número de quadros de *Look Ahead*. A Figura 3.3 ilustra o resultado para os Sinais Tonais. Em todos os casos estudados, o desempenho

do método RTISI-LA com 1 quadro de *Look Ahead* já superou o desempenho do algoritmo G&L.

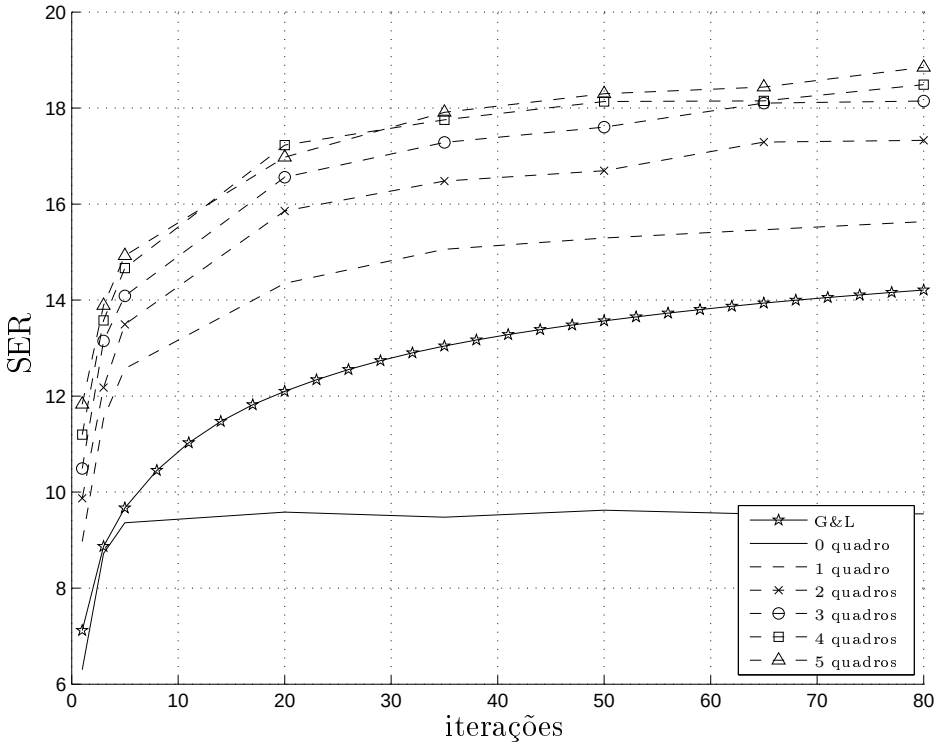

Figura 3.3: Diferentes valores de *Look Ahead* para o Sinal T2 com DFT de 1024 pontos

Outro aspecto a ser ressaltado foi o aumento significativo de desempenho do RTISI-LA para valores de *Look Ahead* de até 3 quadros, não ocorrendo melhora significativa para mais quadros. Estes 3 quadros, por ser usada sobreposição de 75% entre quadros subjacentes, têm sobreposição com o quadro corrente, enquanto os quadros posteriores não. Este é provavelmente o motivo pelo qual o método RTISI-LA não obteve melhoras significativas com a utilização de mais que 3 quadros de *Look Ahead*.

Foi também testado o efeito de se empregar DFTs com 2 e 4 vezes o número L de pontos da janela. Tanto para DFTs com 2048 quanto com 4096 pontos, verificou-se uma grande melhora do desempenho dos métodos RTISI e RTISI-LA com relação ao método G&L. As medidas da SER, para 80 iterações do RTISI-LA, foram de aproximadamente 18 unidades no teste com 1024 pontos e de 29 unidades no teste com 2048 pontos. No entanto, para 4096 pontos não houve melhoras significativas com relação a 2048 pontos, não compensando o custo computacional adicional. Sendo assim, o uso de 2048 pontos pode ser considerado o melhor dos casos testados.

Os casos de 1024 pontos e de 2048 pontos podem ser vistos na Figura 3.4.

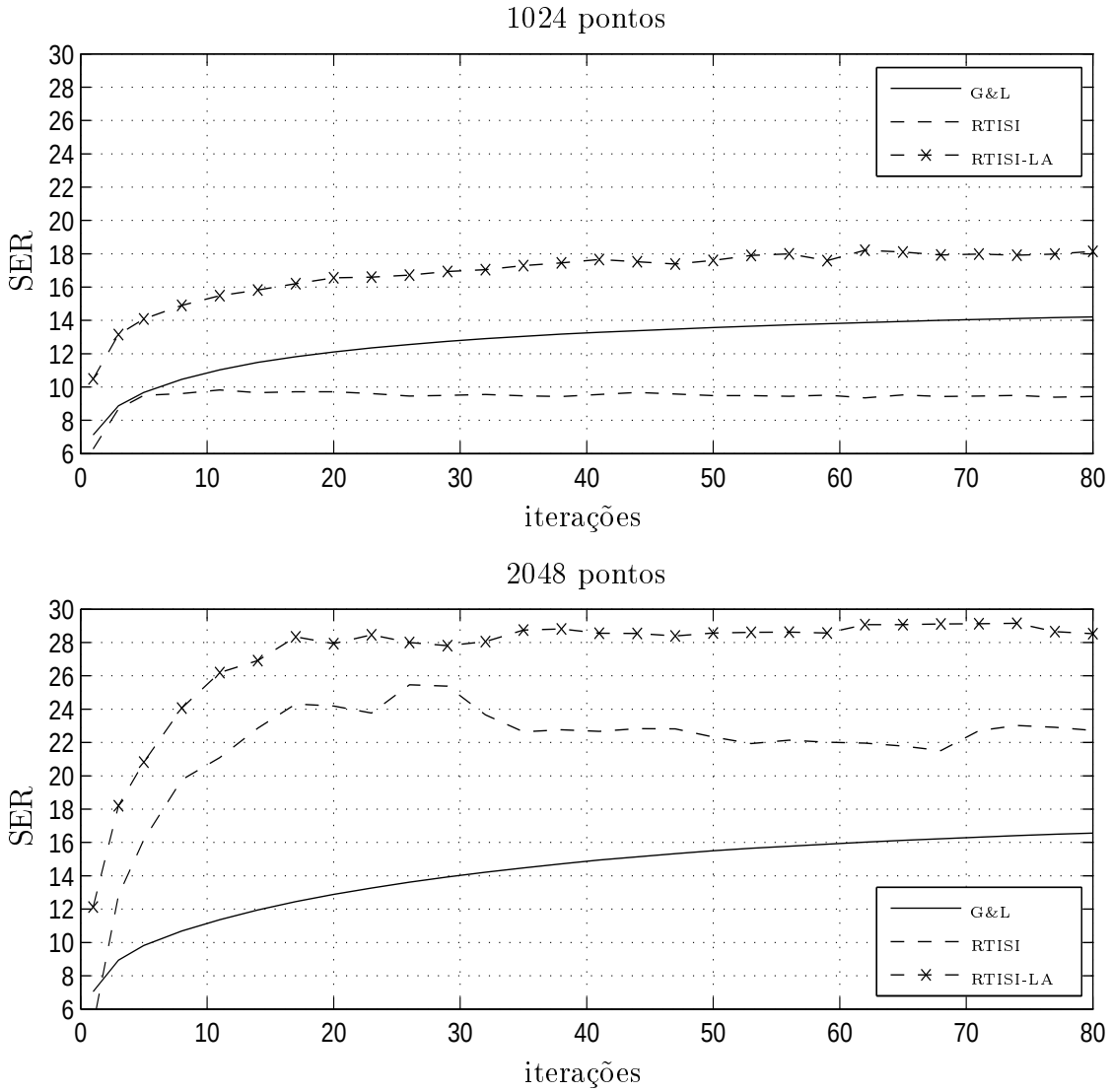

Figura 3.4: Resultados para o Sinal T2 com DFT de 1024 e 2048 pontos

Devido a os algoritmos RTISI e RTISI-LA se restringirem a estimar um quadro por vez, o efeito de *zero-padding*, introduzido ao utilizar mais pontos na DFT do que na janela, é melhor aproveitado por eles através de um efeito de cascata. Para elucidar esta afirmação, pode-se tomar como exemplo a estimação do primeiro quadro no algoritmo RTISI sem *zero-padding*. Neste caso, não há modificações para as L amostras do quadro ao longo das iterações. Isto ocorre porque, como pode ser visto na Equação (2.26), sendo o y_{RTISI} nulo para o primeiro quadro estimado, cada iteração equivalerá ao mesma cálculo: encontrar a transformada inversa de $Y_{w,0}(QS,\omega) = |Y_w(QS,\omega)|$.

Porém, utilizando *zero-padding* na obtenção da MSTFT, a transformada inversa de $Y_{w,0}(QS, \omega)$ conterá, possivelmente, mais que os L pontos não-nulos. Com a aplicação da Equação (2.26), apenas os L primeiros pontos desta transformada inversa serão considerados para tomada da próxima estimativa de fase, devido à multiplicação pela janela, que possui apenas os L primeiros pontos não-nulos. Desta forma, com o aumento do número de iterações, haverá uma tendência de que a energia da transformada inversa se concentre apenas nas L primeiras amostras. À medida que a transformada inversa passar a ter amostras não nulas apenas nas L primeiras amostras, será possível, então, afirmar que a estimativa do quadro será de melhor qualidade, afinal toda a informação da MSTFT do sinal será repassada à estimativa do quadro.

A melhor estimativa do primeiro quadro será utilizada na obtenção do y_{RTISI} da estimação do segundo quadro, como pode ser visto na Equação (2.25). Assim, esse quadro deverá se beneficiar tanto da melhor estimativa do primeiro como também do *zero-padding*, o que deve também levar à obtenção de uma estimativa melhor. O mesmo raciocínio pode ser aplicado aos quadros seguintes, gerando, então, o efeito cascata citado. No caso do RTISI-LA, também se observa esse efeito: as estimativas de cada quadro são beneficiadas, não só pelo uso do *zero-padding*, mas também por partirem de uma melhor estimativa dos quadros anteriores.

Na Seção 2.3.2 foi proposta a utilização de um limiar como critério de parada para o RTISI e o RTISI-LA. Este critério foi testado e seu desempenho comparado ao do critério convencional, do número máximo de iterações. Os resultados obtidos podem ser observados nas Figuras 3.5 e 3.6. Nos testes, foram utilizados 3 quadros de *Look Ahead*. Como, utilizando o critério proposto, o número de iterações realizadas sobre cada quadro pode ser diferente, a comparação foi feita em relação ao número médio de iterações por quadro. Deste modo, pode-se comparar resultados obtidos com o mesmo número total de operações efetuadas para estimar o sinal $x_{\text{final}}[n]$.

Pode-se observar na Figura 3.5 que o uso do limiar no método RTISI gerou resultados aproximadamente iguais aos encontrados usando números fixos de iterações. Pode-se observar também, na Figura 3.6, que o uso do limiar no método RTISI-LA provocou leve queda de desempenho, que, embora menor que a margem de 1 unidade do erro da medida, deve ser levada em consideração por ocorrer consis-

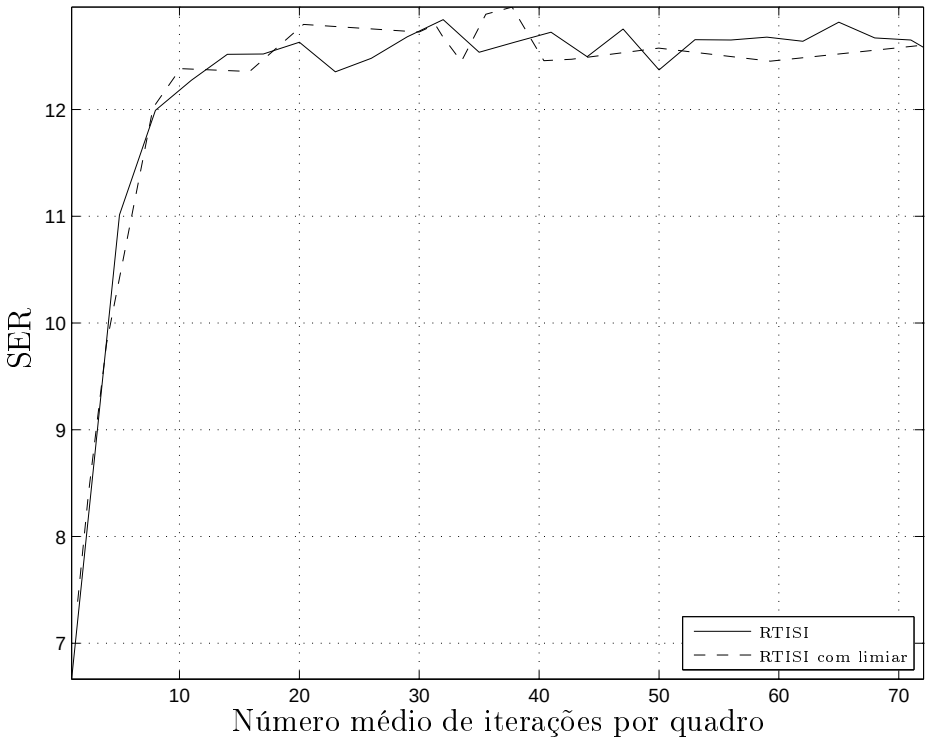

Figura 3.5: Comparação entre RTISI e RTISI com limiar para o Sinal T1 com DFT de 1024 pontos

tentemente em todos os testes. Assim, os métodos RTISI e RTISI-LA modificados para utilizar o limiar não obtiveram a melhora esperada. Esta observação leva ao questionamento de uma das ideias-base da proposta do novo critério: o aumento do número de iterações empregadas sobre um quadro melhora a sua estimativa.

Averiguou-se através dos resultados dos testes que esta ideia é provavelmente falsa. Como visto nas Seções 2.3.1 e 2.4.1, para os métodos RTISI e RTISI-LA, o conjunto dos quadros anteriores ao quadro corrente Q não sofre mudanças durante a estimação deste quadro. Com o aumento do número de iterações, as $L-S$ primeiras amostras do quadro corrente (que têm sobreposição com os quadros anteriores, como ilustrado na Figura 3.7) são forçadas a se aproximar de valores incorretos. Valores que surgem porque os quadros anteriores já são estimativas de quadros reais, tendo, portanto, erros de estimação. Para mostrar essa tendência, foi realizado um teste, descrito a seguir.

Foram tomados os 4 primeiros quadros do sinal original T1, obtidos pelo seu janelamento com a janela de Hamming modificada com duração de 23,2 ms (1024 pontos, para este sinal). Estimou-se, então, o quarto quadro através do método

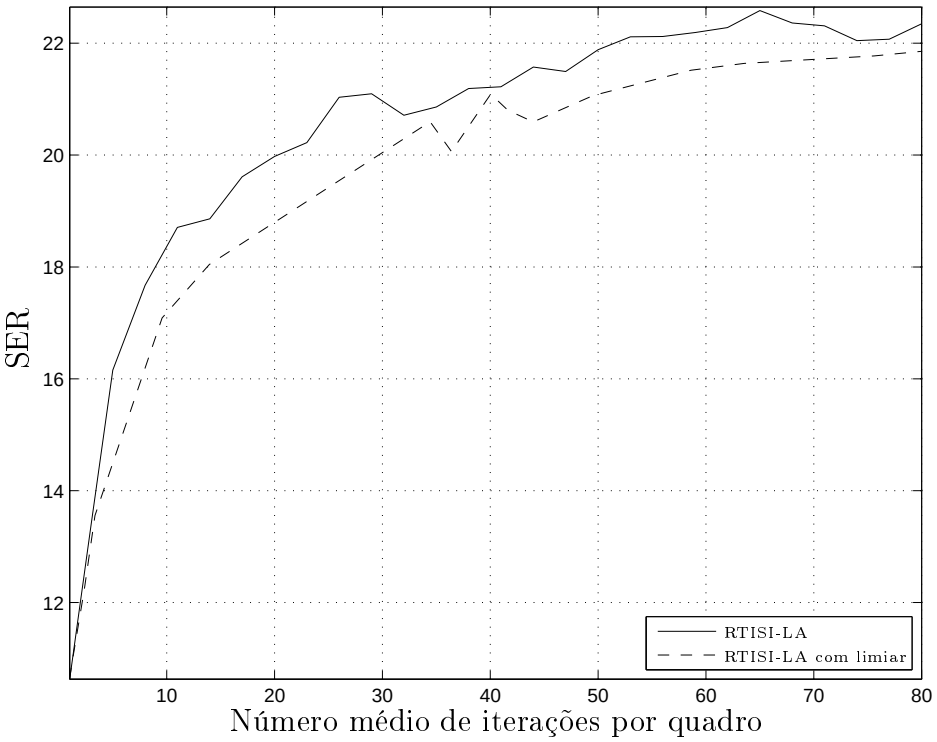

Figura 3.6: Comparação entre RTISI-LA e RTISI-LA com limiar para o Sinal T1 com DFT de 1024 pontos

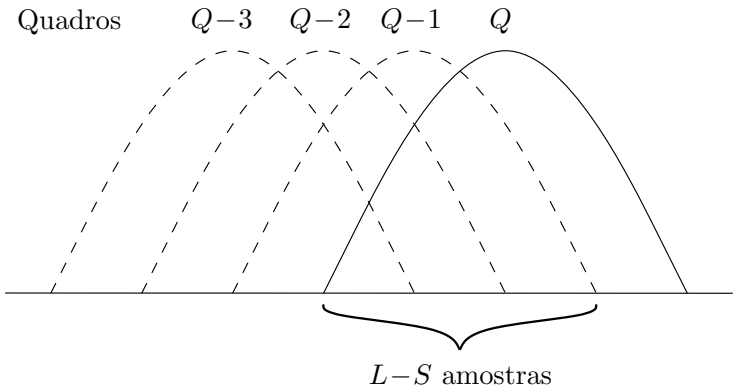

Figura 3.7: Influência dos quadros anteriores no quadro corrente Q

RTISI, utilizando os 3 primeiros quadros originais. Sendo $x_4[n]$ a n-ésima amostra do quarto quadro original e $\hat{x}_4[n]$ a n-ésima amostra da estimativa deste quadro, a Equação (3.6) mostra o erro absoluto em dB, amostra a amostra, obtido na estimação.

$$\text{erro}[n] = 10\log_{10}\left(|x_4[n]-\hat{x}_4[n]|\right). \tag{3.6}$$

Os resultados desse processo iterativo para 0, 50, 150 e 200 iterações podem ser observados na Figura 3.8. O erro absoluto para as primeiras amostras do quadro foi

significativamente menor do que para as demais. No entanto, apesar da disponibilidade dos quadros anteriores que possuem sobreposição com o quadro corrente, o erro teve esta característica para menos que as $L-S$ amostras que constituem a sobreposição. Os 3 quadros anteriores influenciam diretamente, através da Equação (2.26), na estimativa das primeiras 256 amostras do quadro corrente. De modo semelhante, há a influência direta de 2 quadros anteriores na estimativa desde a amostra 257 até a amostra 512 e de 1 quadro anterior desde amostra 513 até a amostra 768, o que pode ser observado visualizado pela Figura 3.7. As demais amostras do quadro corrente não são influenciadas diretamente por nenhum quadro anterior. Essa queda da influência dos quadros anteriores é a provável razão para o erro aumentar ao longo do quadro. Deve-se ressaltar também que, mesmo com a utilização de uma janela que possui maior energia no centro do quadro, o erro manteve a tendência crescente ao longo das amostras deste.

Neste teste, como os quadros anteriores foram provenientes do sinal original, a maior influência deles sobre o quadro atual constituiu um menor erro de estimação. No caso em que os quadros anteriores são estimativas dos quadros do sinal, essa influência vicia a estimativa do quadro atual porque aproxima as amostras do quadro corrente de valores, em geral, incorretos.

Assim, dadas estimativas dos quadros anteriores, o quadro corrente passa a ter uma estimativa pior. Tem-se, portanto, uma composição de erros com o aumento das iterações, o que não possibilita a melhora pretendida com utilização do critério proposto.

A leve queda de desempenho no caso do RTISI-LA com limiar pode ser uma outra consequência da aproximação forçada das $L - S$ primeiras amostras para os valores incorretos. Ao aproximar essas amostras desses valores, a vantagem adquirida ao usar quadros de *Look Ahead* é reduzida. Isso, então, desfavorece a utilização do critério por limiar.

3.2.1.2 Conclusão

Para os Sinais Tonais, concluiu-se que o algoritmo RTISI-LA consegue desempenho bastante superior aos demais métodos testados. Um outro aspecto que pode ser destacado refere-se ao número de pontos utilizados para o cálculo da MSTFT dos

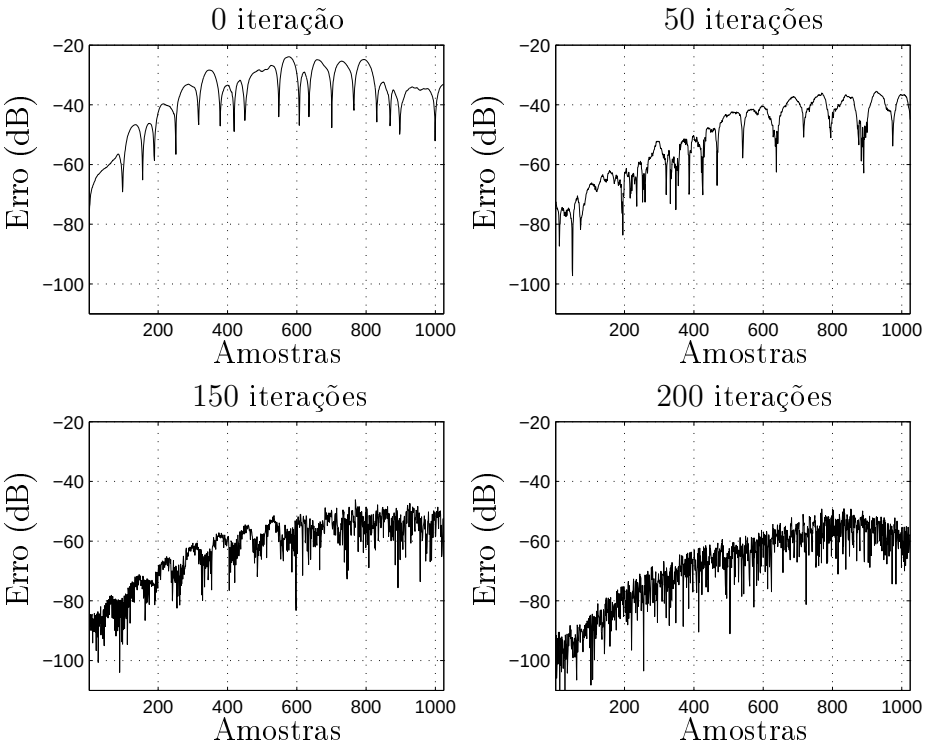

Figura 3.8: Erro absoluto amostra a amostra na estimação do quarto quadro do sinal T1

sinais. Dentre as configurações testadas para a reconstrução, os melhores resultados foram obtidos ao se utilizar o dobro do número de pontos da janela.

Por fim, verificou-se também que a utilização de mais que 3 quadros de *Look Ahead* não levou a ganhos de desempenho significativos.

Quanto à medição da SER, os valores medidos com $N = 2^{14}$ foram bastante próximos aos medidos empregando-se N igual a duas vezes ou 4 vezes o número de pontos da janela. Isto leva a crer que, em situações práticas, utilizar $N = 2L$ ou $N = 4L$ levará a valores bem próximos dos valores reais da medida.

3.2.2 Sinais Percussivos

3.2.2.1 Resultados

Para sinais percussivos, a delimitação temporal dos eventos que ocorrem no sinal costuma ser mais importante que para sinais tonais. Sinais percussivos possuem eventos de curta duração, bem localizados no tempo. Como essa delimitação temporal está fortemente relacionada com a fase do sinal, sinais desse tipo constituem um grande desafio para o desempenho dos métodos de reconstrução de fase.

Uma estimativa de fase mal feita pode causar espalhamento temporal de eventos que deveriam ser bem localizados no tempo, prejudicando a qualidade do áudio.

Foram utilizados 3 sinais percussivos nos testes, descritos na Tabela 3.2.

Tabela 3.2: Descrição dos Sinais Percussivos

Código	Descrição
P1	Trecho de solo de bateria de Joe Morello em estilo jazzístico. Soam tambores, caixa e chimbau.
P2	Outro trecho do mesmo solo, com rápida sucessão de notas. Soam bumbo, caixa e pratos, os pratos predominando no final.
P3	Trecho de solo de castanholas, extraído do CD *Perceptual Audio Coders* editado pela *Audio Engineering Society*. Sinal caracterizado por sons com ataque bem definido e curta duração.

Dentre estes sinais, P1 e P2 apresentaram resultados similares, comentados a seguir. Os resultados de P3, que apresentam diversas peculiaridades, são comentados no fim da seção.

O comportamento dos métodos para o sinal P1 é ilustrado na Figura 3.9 tanto para testes utilizando DFTs de 1024 pontos nas representações de cada quadro de X_w e Y_w quanto para DFTs de 2048 pontos. Os gráficos obtidos para o sinal P2 são omitidos por serem semelhantes aos gráficos do sinal P1. Para DFTs de 4096 pontos, os resultados foram análogos aos obtidos para as de 2048 pontos.

Pôde-se notar que os resultados para DFTs de 2048 pontos foram melhores que os obtidos para 1024 pontos, tal como observado para os Sinais Tonais. Desta forma, a comparação entre os métodos para esses sinais corrobora os comentários feitos para Sinais Tonais.

Nos testes relativos ao número de quadros de *Look Ahead*, os sinais P1 e P2 obtiveram melhoras significativas até 5 quadros, como pode ser visto nas Figuras 3.10 e 3.11. Este valor, maior que o do melhor valor encontrado para Sinais Tonais, abrange 2 quadros que não tem sobreposição com quadro corrente (para a

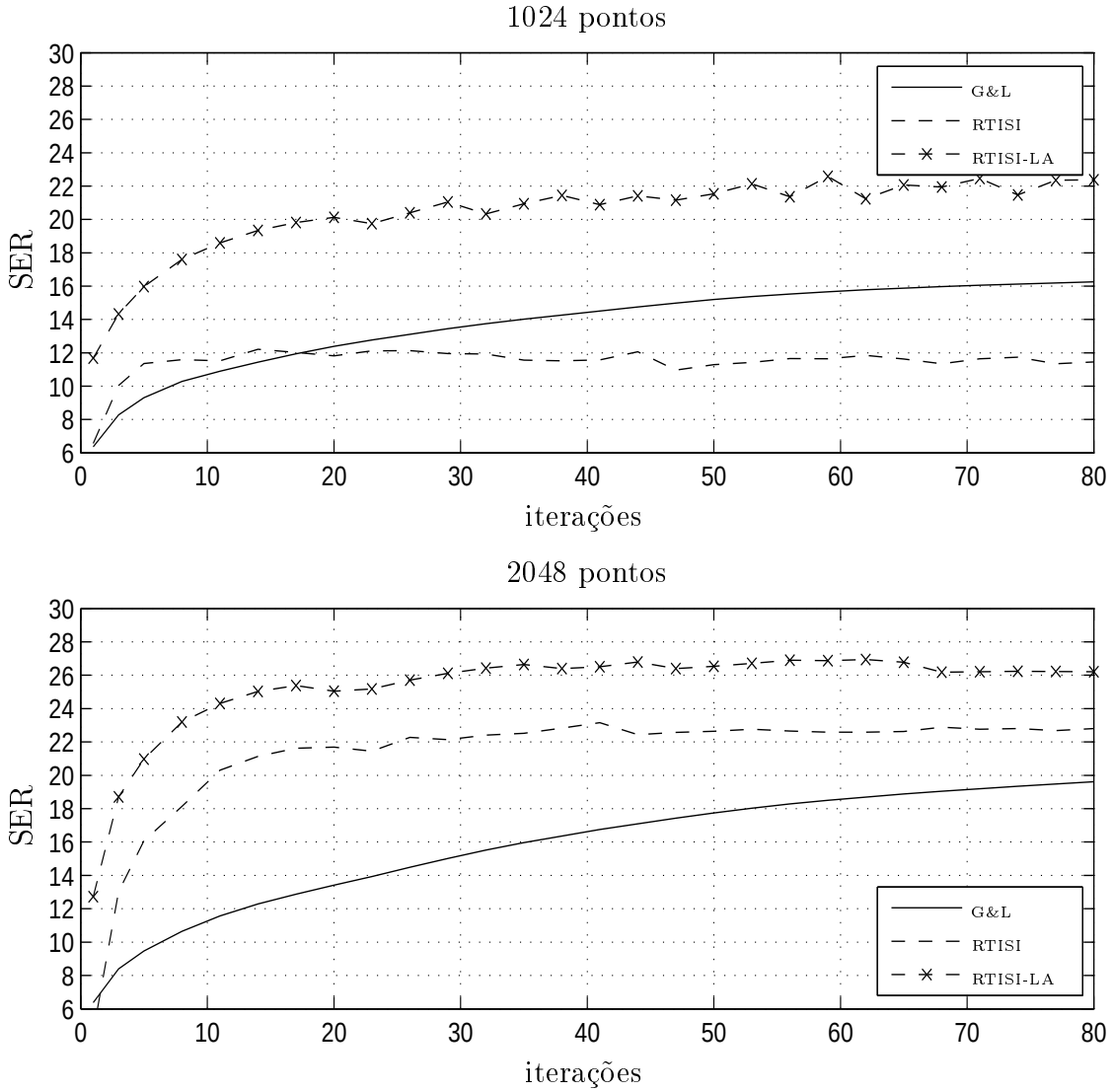

Figura 3.9: Resultados para o Sinal P1 com DFT de 1024 e 2048 pontos

sobreposição de 75% entre quadros adjacentes utilizada). Estes 2 últimos quadros do bloco de *Look Ahead* possivelmente ajudam a mitigar os erros das estimativas parciais dos 3 quadros de *Look Ahead* que têm sobreposição com o quadro corrente. Isto pode ser mais relevante para sinais percussivos devido à ocorrência de eventos de curta duração.

Devido a essa pequena duração dos eventos, é possível que alguns dos quadros de *Look Ahead* tenham energia muito maior que a do quadro corrente. Com isso, valores estimados para as amostras do quadro corrente que também pertençam aos quadros de *Look Ahead* com maior energia deverão sofrer grande influência das estimativas parciais desses quadros. Com mais quadros de *Look Ahead*, deverão ser

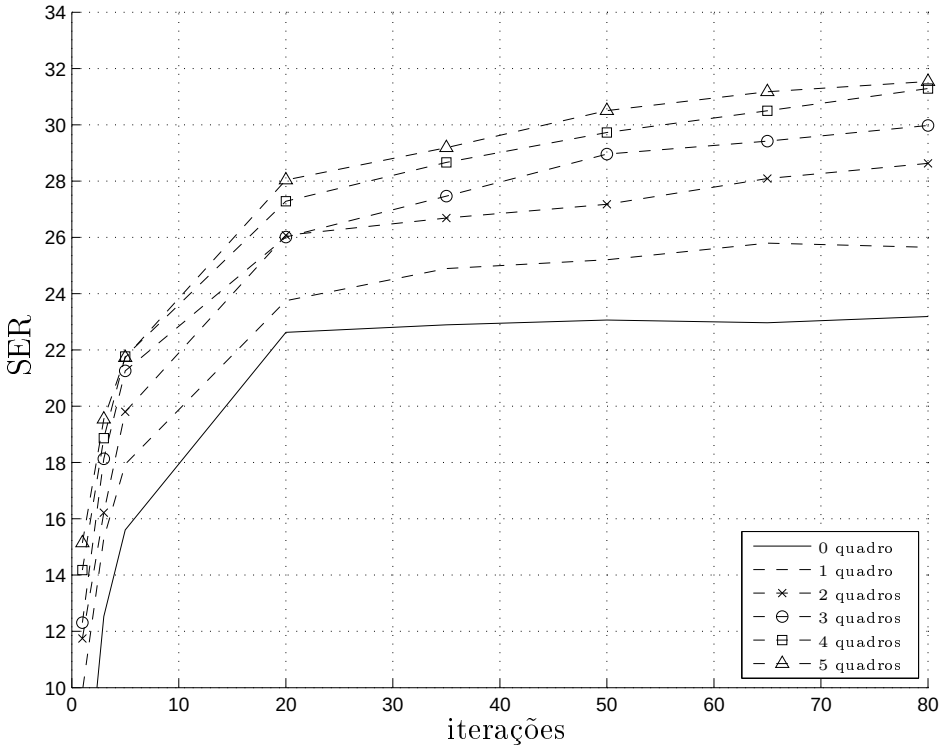

Figura 3.10: Diferentes valores de *Look Ahead* para o Sinal P2 com DFT de 2048 pontos

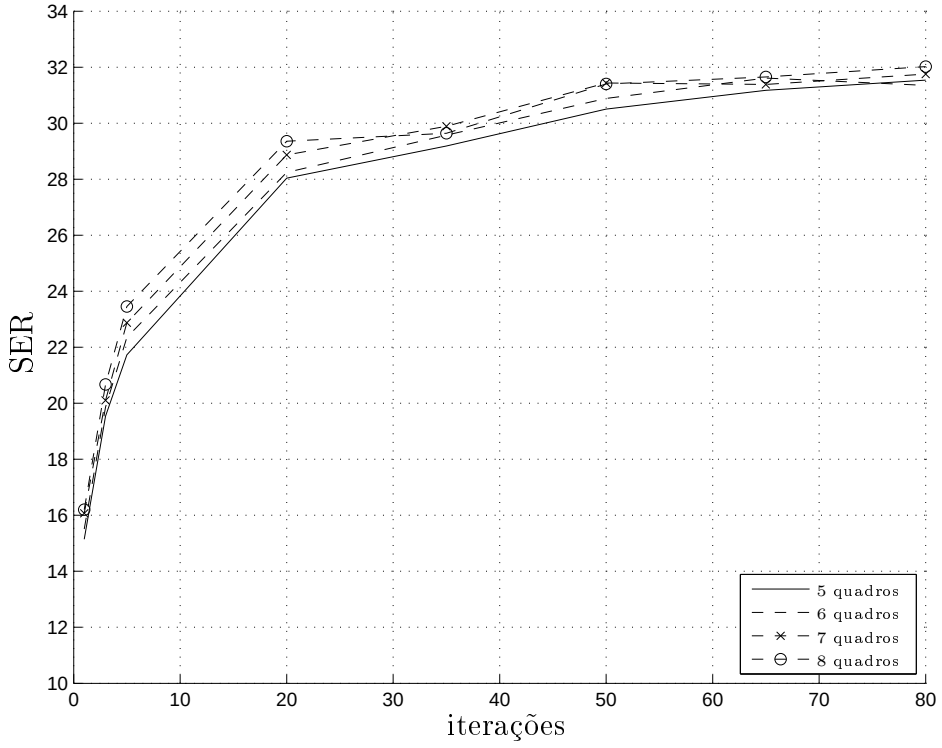

Figura 3.11: 5 a 8 quadros de *Look Ahead* para o Sinal P2 com DFT de 2048 pontos

obtidas melhores estimativas parciais, favorecendo a obtenção de melhores valores para essas amostras. No método RTISI-LA, quando se termina a estimação de um quadro, este é tomado como pronto, não sendo mais possível modificar sua estimativa. Passa-se, então, à estimação do próximo (até então, um dos quadros de *Look Ahead*). Como já citado na Seção 3.2.1.1, as primeiras $L - S$ amostras do quadro corrente devem ser forçadas, com o aumento do número de iterações, a se aproximar de valores possivelmente incorretos calculados com base em y_{RTISI}. Tendo sido os valores destas $L-S$ amostras mais bem estimados nas estimativas dos quadros anteriores, estas deverão, por fim, ser melhor estimadas no quadro atual. Este efeito cascata é semelhante àquele discutido na Seção 3.2.1.1 com relação à utilização de *zero-padding*.

Dessa forma, a melhor estimativa parcial dos quadros de maior energia, durante a estimativa do quadro corrente, deve melhorar também a qualidade de suas estimativas finais, possibilitando obter maior precisão na delimitação temporal de eventos.

Os testes para os Sinais P1 e P2 com os algoritmos RTISI utilizando critério de parada por limiar também apresentaram resultados próximos aos obtidos com critério de parada por número máximo de iterações, como ocorreu para os Sinais Tonais. A Figura 3.12 ilustra esses resultados para o método RTISI e a Figura 3.13 para o método RTISI-LA.

Para o sinal P3, o método G&L e o método RTISI-LA tiveram resultados bastante superiores aos do método RTISI, tanto para 1024 quanto para 2048 e 4096 pontos. Isso possivelmente ocorreu porque este sinal é pouco estacionário, como pode ser observado na Figura 3.14.

No tratamento de sinais pouco estacionários, métodos que utilizam a informação de todos os quadros que se sobrepõem ao quadro corrente devem ter um melhor desempenho, por terem mais chances de modelar as rápidas variações das propriedades estatísticas do sinal. Ao se utilizar apenas a informação de quadros passados e do quadro corrente, como no caso do RTISI, as características estatísticas desses quadros terão maior influência sobre a estimativa do quadro corrente, que assim não 'se preparará' para as variações que devem ocorrer nos próximos quadros.

Os resultados da comparação entre os métodos são ilustrados na Figura 3.15,

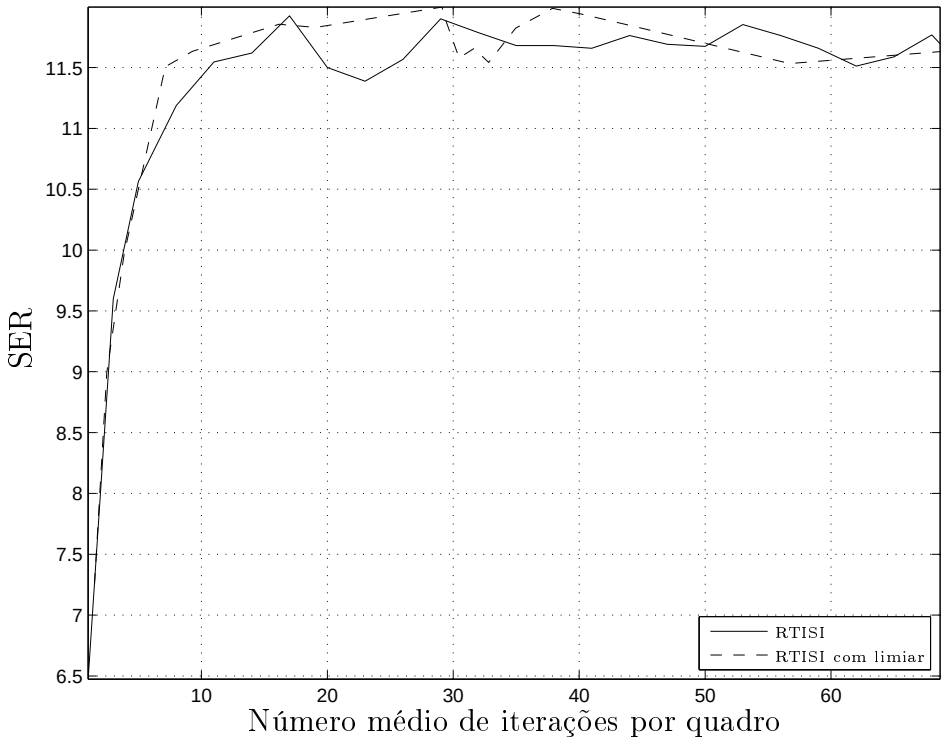

Figura 3.12: Comparação entre RTISI e RTISI com limiar para o Sinal P2 com DFT de 1024 pontos

onde o método RTISI corresponde ao caso particular do RTISI-LA sem quadros de *Look Ahead*.

O método G&L só é efetivamente superado pelo método RTISI-LA quando são utilizados 3 quadros de *Look Ahead*. Isto denota que o RTISI-LA só conseguiu ter melhor desempenho ao associar, pelo menos, toda a informação da vizinhança do quadro corrente, isto é, a informação de todos os quadros que se sobrepõem a esse quadro.

Foi possível verificar também melhora no desempenho do método RTISI para até 8 quadros de *Look Ahead*, como mostrado na Figura 3.16. Não foram, porém, efetuados testes para mais quadros, uma vez que que seriam extremamente custosos computacionalmente. Este estudo também seria de uma situação extremamente particular, o que foge ao objetivo do trabalho de verificar os melhores parâmetros para a utilização dos métodos em situações práticas.

Nos testes para o sinal P3 utilizando limiar como critério de parada, os métodos RTISI e RTISI-LA mostraram desempenho superior ao obtido com critério de parada por número máximo de iterações. Isto ocorreu para 1024, 2048 e 4096

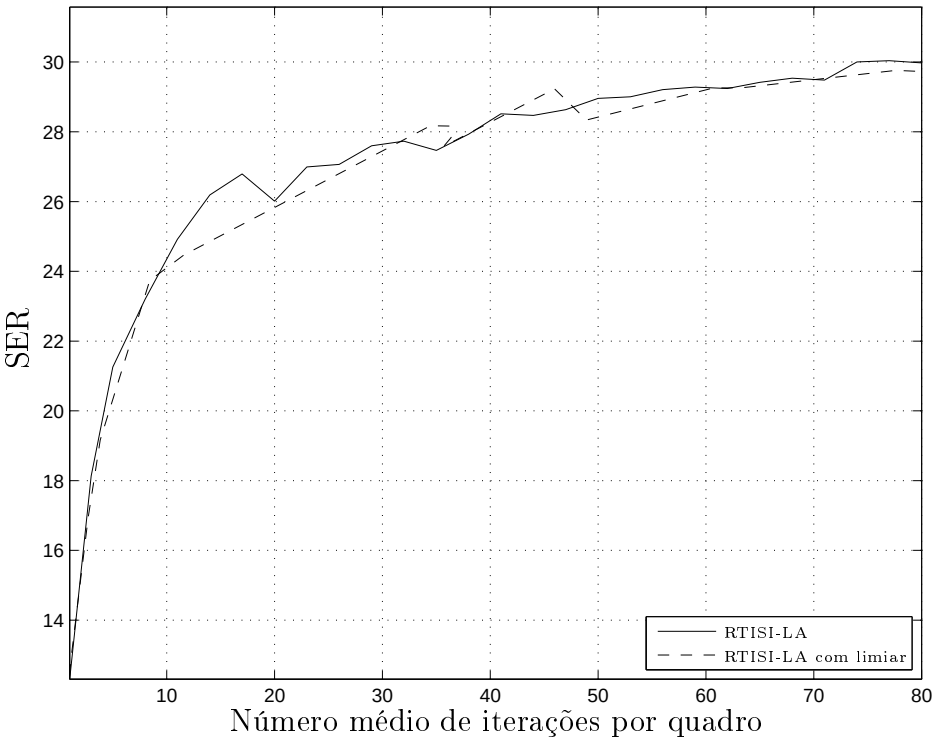

Figura 3.13: Comparação entre RTISI-LA e RTISI-LA com limiar para o Sinal P2 com DFT de 2048 pontos

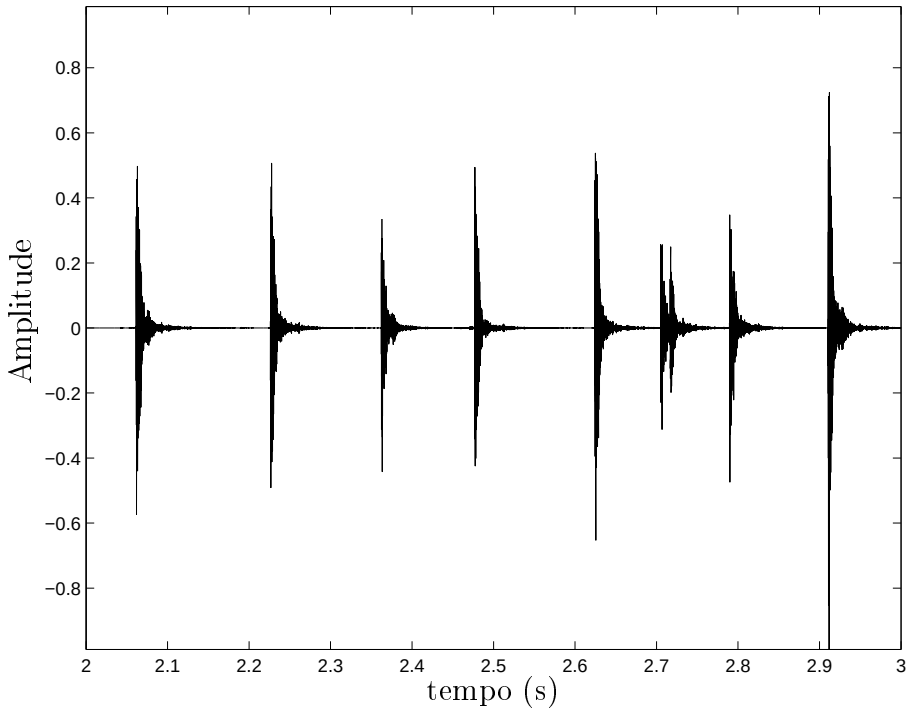

Figura 3.14: Forma de onda de trecho de 2s a 3s do Sinal P3 original

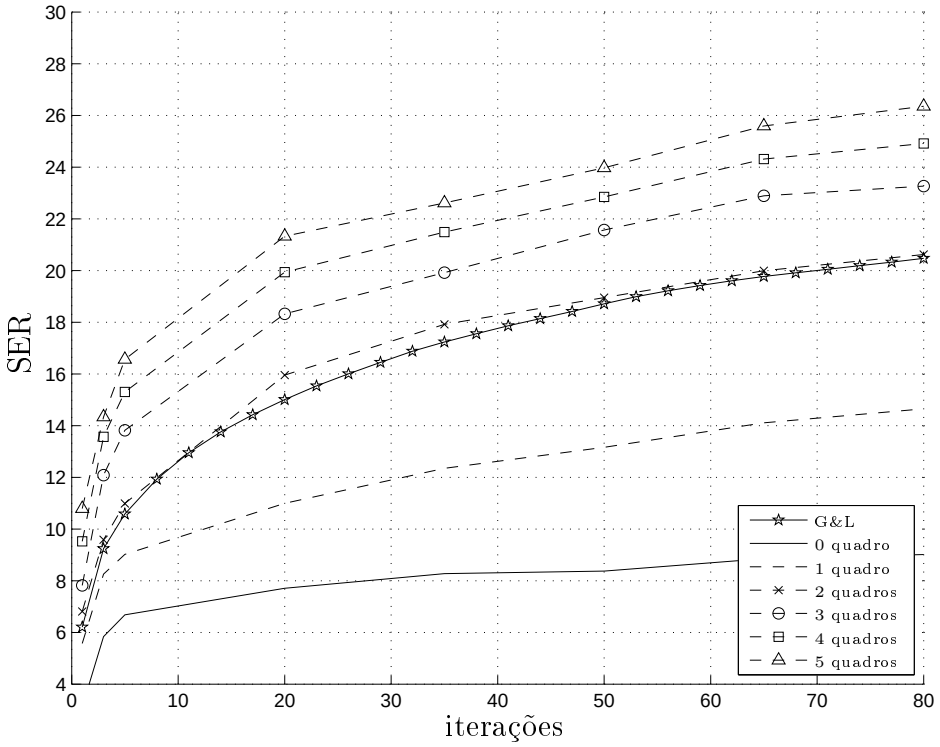

Figura 3.15: Diferentes valores de *Look Ahead* para o Sinal P3 com DFT de 2048 pontos

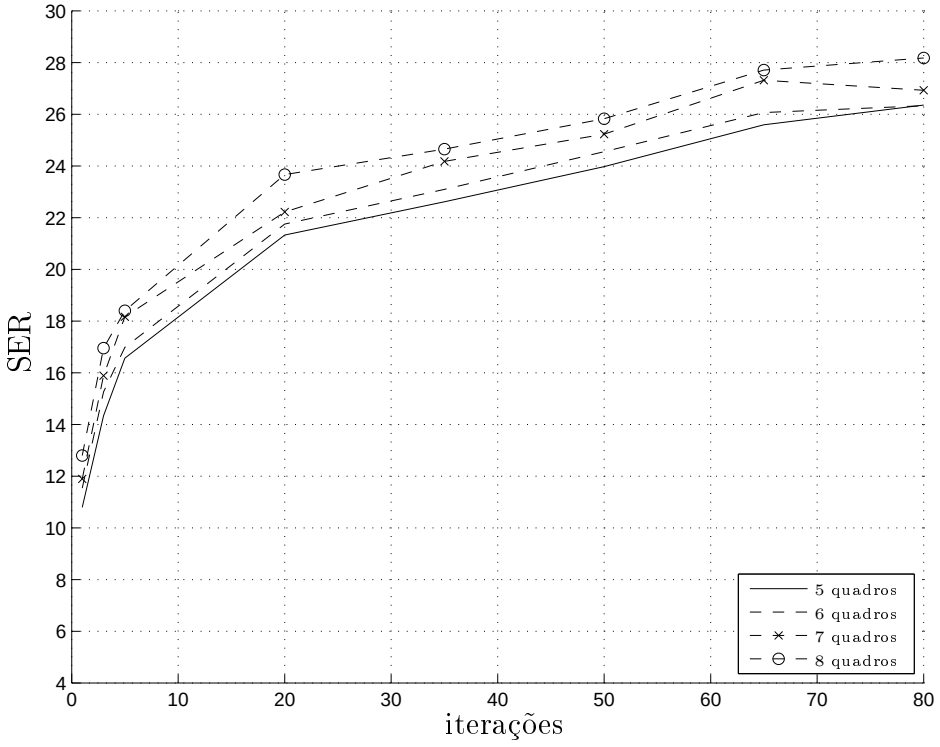

Figura 3.16: 5 a 8 quadros de *Look Ahead* para o Sinal P3 com DFT de 2048 pontos

pontos. Este desempenho superior, mostrado para o RTISI-LA na Figura 3.17, é justificável pelo fato do sinal conter muitas áreas de silêncio, nas quais o uso do

critério de parada tradicional gera iterações desnecessárias. Já o método com limiar gera mais iterações apenas nos trechos de maior interesse.

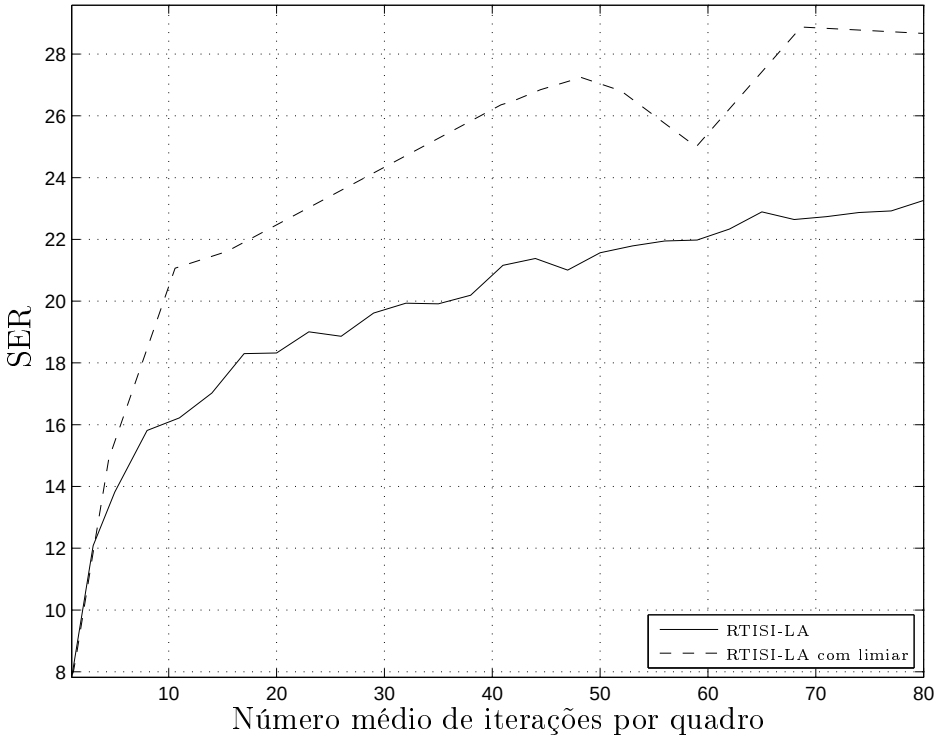

Figura 3.17: Comparação entre RTISI-LA e RTISI-LA com limiar para o Sinal P3 com DFT de 2048 pontos

O problema da composição de erros do algoritmo por limiar, discutido na Seção 3.2.1.1, não ocorre de forma determinante para este sinal. Esse sinal é caracterizado por grandes zonas de silêncio e ataques com rápido decaimento, o que é mostrado na Figura 3.14. Nas zonas de silêncio, os quadros têm baixa energia, e os erros gerados na estimação de suas fases não interferem de forma significativa na estimação dos quadros com maior energia. E, não havendo muitos quadros de maior energia adjacentes, não haverá muita propagação de erro entre esses quadros.

3.2.2.2 Conclusão

Quanto aos métodos, o RTISI-LA apresentou resultados superiores aos demais.

Com relação ao número de pontos da DFT nas representações de X_w e Y_w, a utilização de $N = 2L$ se mostrou, novamente, a escolha mais eficiente.

No que se refere ao melhor número de quadros de *Look Ahead*, os resultados para 5 quadros foram bastante próximos àqueles obtidos para mais quadros. Como o aumento do número de quadros implica aumento do custo computacional, este valor é indicado para sinais percussivos.

Os resultados para os Sinais Tonais e para os Sinais Percussivos tiveram dois pontos em comum. Para as duas famílias, o desempenho relativo dos métodos e sua variação com o número de pontos da DFT foi similar. Assim sendo, a fim de tornar o texto das próximas seções mais claro e sucinto, resultados quanto a estes aspectos só serão nelas citados caso haja alguma diferença importante a ser ressaltada.

Para um dos sinais testados, o critério de parada por limiar foi mais eficaz. Deve-se ressaltar, porém, que como os resultados deste sinal apresentavam muitas peculiaridades, não foram considerados representativos de uma análise geral do comportamento dos métodos.

Ainda, devido a os métodos com o critério de parada por limiar não terem apresentado melhora em relação à parada por número máximo de iterações nem para os Sinais Tonais nem para os Sinais Percussivos, testes com esse critério de parada não são reportados nas seções seguintes.

3.2.3 Sinais Tonais e Percussivos

3.2.3.1 Resultados

Os sinais descritos a seguir possuem características mistas dos sinais analisados até este ponto. São registros de instrumentos tonais e percussivos soando simultaneamente.

Dentre os testes realizados para os Sinais Tonais e Percussivos, cabe apenas ressaltar os resultados encontrados para o Sinal TP1. Como este sinal foi um dos sinais utilizados para testes por Beauregard *et al.* [4], seus resultados também servem para validar as implementações realizadas. Tanto TP1 quanto os demais sinais apresentaram comportamento semelhante àqueles analisados nas seções anteriores. Porém, é relevante mostrar os resultados para TP1 porque este sinal tem taxa de amostragem diferente dos demais.

Como TP1 possui taxa de amostragem de 22050Hz, o teste voltado para descobrir o melhor número de pontos N a serem utilizados na obtenção de Y_w pode

Tabela 3.3: Descrição dos Sinais Tonais e Percussivos

Código	Descrição
TP1	Trecho de música em estilo *pop* dançante, com guitarra, teclados, bateria eletrônica, baixo elétrico e voz.
TP2	Trecho do final do Concerto para Orquestra de Béla Bartók, com forte presença de metais.
TP3	Duo de piano e bateria, em estilo jazzístico, executados respectivamente por Dave Brubeck e Joe Morello.
TP4	Trecho instrumental da música *Pieces*, de Patricia Barber, do álbum *Verse*. Soam guitarra-base (com efeitos de atraso), guitarra-solo e bateria.
TP5	Trecho da música *Take Five*, tocada por *The Dave Brubeck Quartet*. Soam piano, baixo e bateria acompanhando solo de saxofone alto.

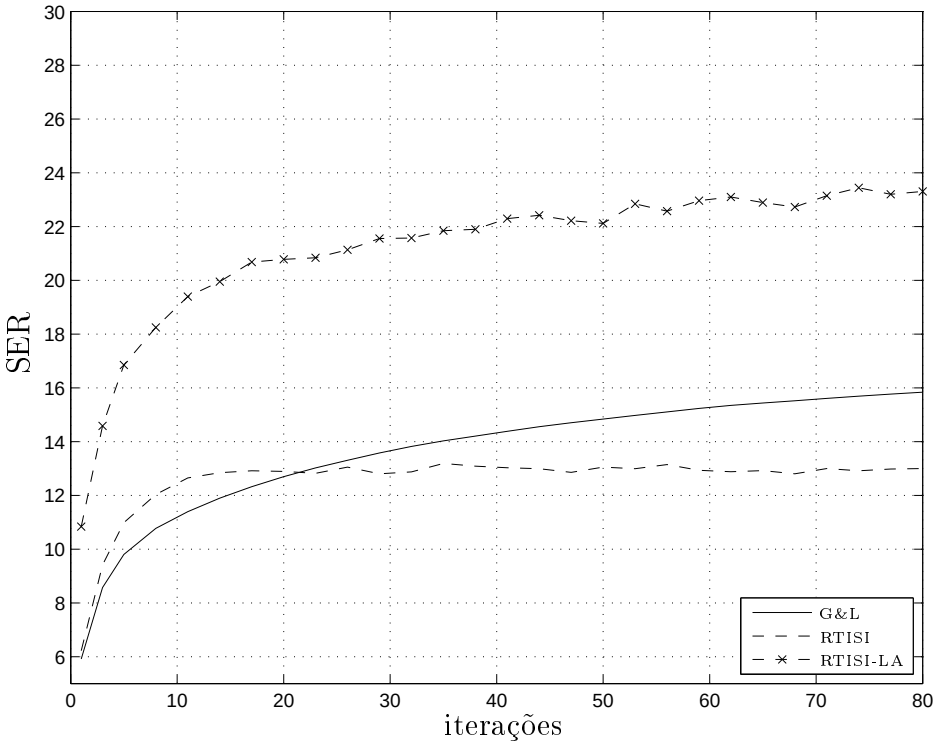

Figura 3.18: Resultados para o Sinal TP1 com DFT de 512 pontos

ser realizado em outra situação.

Os testes para comparação entre os métodos, mostrados nas Figura 3.18 para $N=512$ e na Figura 3.19 para $N=1024$, tiveram resultados semelhantes aos obtidos com sinais puramente tonais ou puramente percussivos: o método RTISI-LA melhor em todos os casos; o RTISI melhor que o G&L apenas para poucas iterações com $N=L$, e sempre melhor com $N=2L$. Os resultados com $N=2048$ são semelhantes àqueles obtidos com $N=1024$, sendo por isso o seu gráfico omitido.

Ao variar o número de pontos utilizado em Y_w, a escolha de $N=2L$ ($N=1024$, neste caso) mostrou-se, mais uma vez, a mais indicada. O resultado para $N=1024$ pontos é mostrado na Figura 3.19, tendo sido omitido o resultado similar obtido para 2048 pontos. Esse resultado deve ser ressaltado porque sugere que a utilização de $N=2L$ pode aprimorar o desempenho da reconstrução independentemente da taxa de amostragem utilizada.

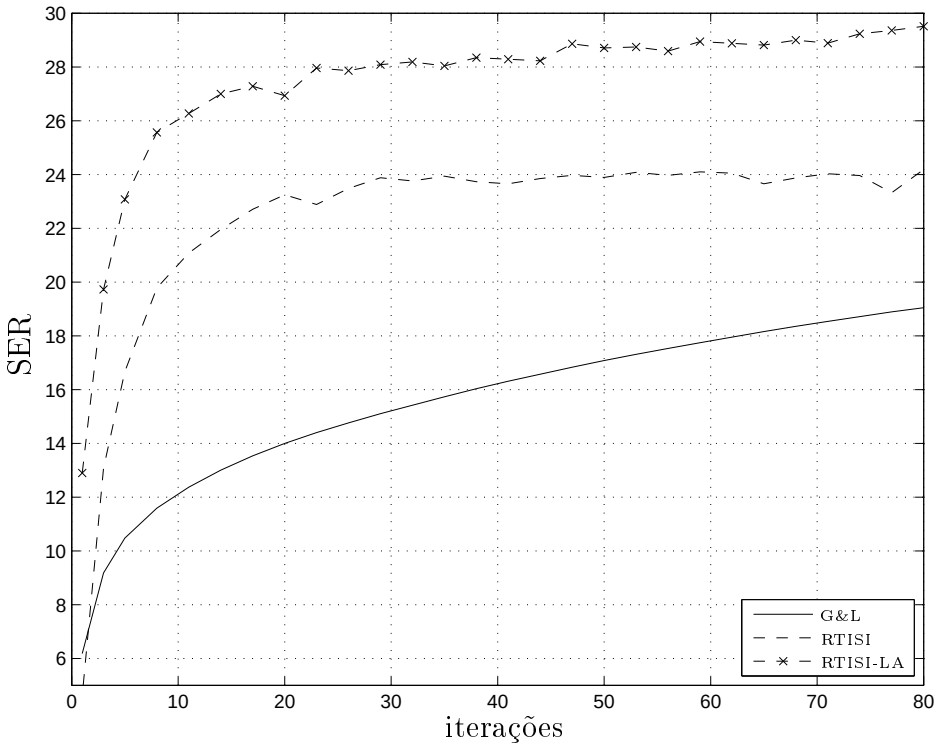

Figura 3.19: Resultados para o Sinal TP1 com DFT de 1024 pontos

3.2.3.2 Conclusão

Devido às características mistas desses sinais, era esperado que, para os Sinais Tonais e Percussivos, os métodos tivessem comportamentos próximos aos descritos

nas Seções 3.2.1.1 e 3.2.2.1. Isto foi confirmado pelos resultados obtidos.

Mesmo para uma taxa de amostragem diferente, 22050 Hz, a utilização de $N = 2L$ mostrou-se a mais eficiente, indicando a robustez desta parametrização.

3.2.4 Sinais de Fontes Pré Mixagem

Nesta seção são apresentados os resultados relativos à testes com o algoritmo MISI. Para todos os sinais foi efetuada a reconstrução com todos os métodos citados neste trabalho, possibilitando comparações entre os resultados de todos esses métodos.

3.2.4.1 Resultados

Os testes foram realizados utilizando trechos das faixas intituladas *Roads*, *Que Pena Tanto Faz* e *Remember The Name*. A faixa *Roads* deve ser creditada à dupla *Bearlin (Ignasi Calvo e Jordi Rabascall)*, tendo sido produzida por *Sergi Vila e Bearlin*. O *site* da dupla é *www.bearlin.net*. A divulgação destes créditos constitui exigência da licença de utilização da faixa para fins científicos. A faixa *Que Pena Tanto Faz* pode ser utilizada sob os termos da licença *Creative Commons Attribution Noncommercial 3.0* (*http://creativecommons.org/licenses/by-nc/3.0/br/*) e deve ser creditada a *Tamy* e *Curvemusic*, tendo sido obtida inicialmente em *http://ccmixter.org/curve/view/contest/sources*. A faixa *Remember The Name* pode ser utilizada sob os termos da licença *Creative Commons Attribution-NonCommercial 2.5* (*http://creativecommons.org/licenses/by-nc/2.5/*), e deve ser creditada a *Fort Minor*, *Warner Bros. Records* e *Machine Shop Recordings*, sendo *http://www.myspace.com/fortminor* o *site* do autor. O sinal de mistura obtido junto com o sinal das fontes foi mixado por *Michel Desnoues*, pertencente à *Telecom ParisTech*.

Os testes com o algoritmo MISI foram realizados utilizando registros de fontes presentes nas faixas, mas ainda não mixadas. O aumento do número de fontes foi realizado gradativamente de forma a analisar sua influência no desempenho do método. Os sinais das fontes de cada faixa foram misturados apenas com sinais da mesma faixa. O acréscimo foi feito na ordem em que as faixas são mostradas na Tabela 3.4, na Tabela 3.5 e na Tabela 3.6.

Tabela 3.4: Descrição dos Sinais das Fontes de *Roads*

Código	Descrição
BR1	Acorde tocado com batida em violão com cordas de aço com o auxílio de uma palheta.
BR2	Soam notas de baixo, tocadas lentamente.
BR3	Soa um acorde sintetizado.
BR4	Voz masculina cantada com eco.
BR5	Soam acordes de piano, com algumas notas tocadas individualmente.
BR6	Levada lenta de bateria, com chimbau, bumbo uma rápida batida no prato.

Tabela 3.5: Descrição dos Sinais das Fontes de *Que Pena Tanto Faz*

Código	Descrição
QP1	Soam acordes de violão tocado com batida de bossa.
QP2	Voz cantada feminina.

Tabela 3.6: Descrição dos Sinais das Fontes de *Remember The Name*

Código	Descrição
RN1	Síntese de violino soando notas isoladas acompanhado por acordes.
RN2	Soam notas de baixo bastante graves.
RN3	Duas vozes masculinas cantando em estilo *Rap*.
RN4	Bateria eletrônica com batida forte.

Para todos os sinais foram realizados dois tipos de testes: no primeiro, foi avaliada a qualidade da reconstrução dos sinais de acordo com o número de fontes incluídas; no segundo, o método MISI foi comparado com os demais métodos.

A Figura 3.20 ilustra os resultados do primeiro teste. O método MISI, como mostrado na Seção 2.5, utiliza o sinal da mistura para auxiliar na estimação da fase dos sinais das fontes. Com o aumento do número de fontes, espera-se que a reconstrução dos sinais piore. Isto deve ocorrer porque com este aumento, diminui a porcentagem de energia no sinal devida a cada fonte independente. Assim, cada fonte passa a ter menor influência no cálculo do erro, aproximando o comportamento do método MISI do comportamento do método G&L.

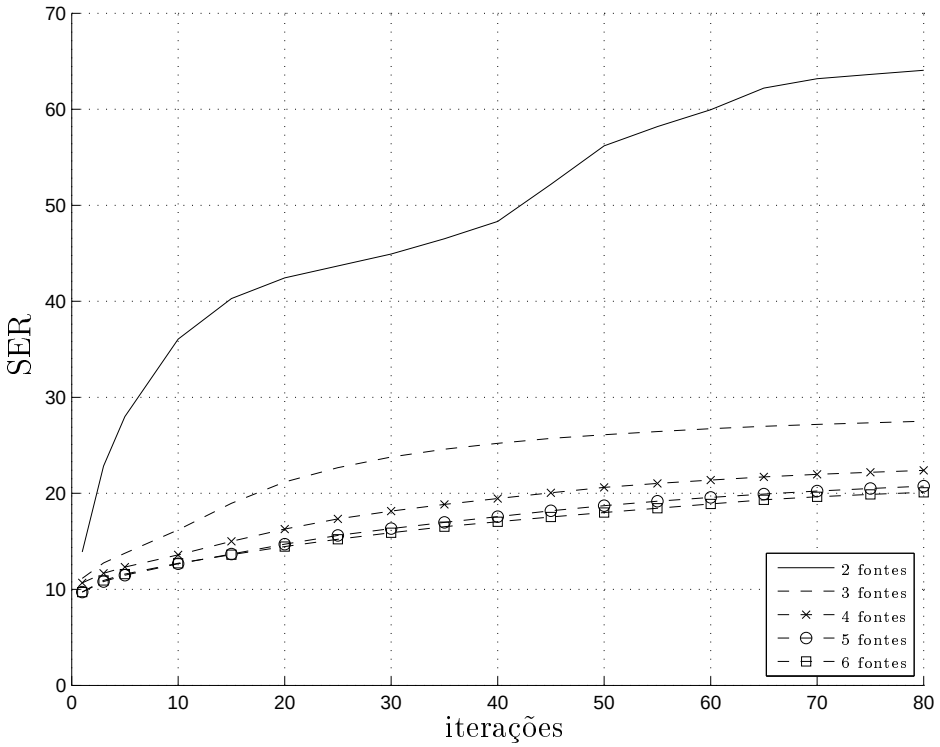

Figura 3.20: Resultados MISI para diferentes números de fontes para o Sinal BR1 com DFT de 1024 pontos

No segundo tipo de teste, onde o MISI foi comparado com os demais métodos, as comparações para cada sinal de teste foram realizadas considerando duas opções: a situação mais favorável para o método MISI, usando o sinal da mistura do menor número de fontes que contenha o sinal de teste; e a situação menos favorável, usando o sinal da mistura de todas as fontes disponíveis para a faixa desse sinal de teste. É importante notar que a comparação entre diferentes números de fontes não deve

ser feita de maneira quantitativa, apenas qualitativa. Espera-se que o acréscimo de uma fonte à mistura piore o desempenho do método. No entanto, os desempenhos do método para misturas diferentes com o mesmo número de fontes não podem ser comparados, devido a os resultados serem dependentes das características das fontes.

Para o Sinal QP2, o método MISI utilizado com a mistura de 2 fontes obteve resultado bem melhor que os demais métodos, como pode ser visto na Figura 3.21. Neste caso, ao se utilizar o sinal da mistura contendo duas fontes, o método consegue aproveitar o sinal de erro suficientemente para aprimorar a estimativa do sinal temporal de cada fonte. Assim, a realimentação e a inicialização feita com a fase da mistura representaram uma vantagem em relação aos demais métodos, que partiram de fase nula.

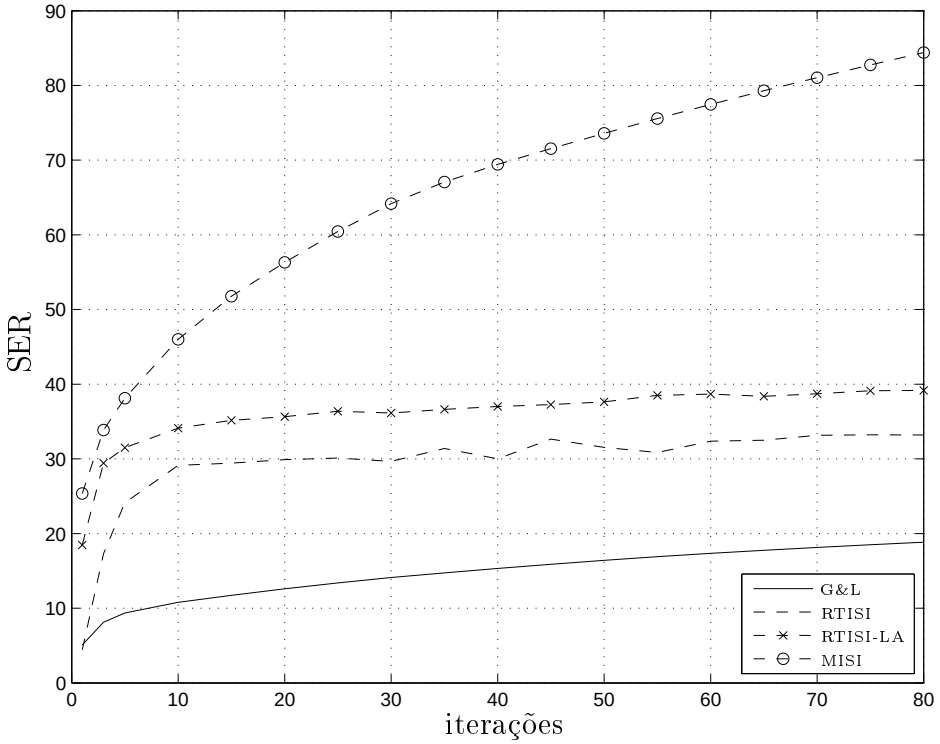

Figura 3.21: Resultados para o Sinal QP2 com DFT de 2048 pontos

O comportamento superior do método MISI mostrado na Figura 3.21 para a mistura de 2 fontes também ocorre para os demais sinais testados para esta situação. Porém, em alguns casos, com o aumento do número de fontes, o método foi superado pelo RTISI-LA (como na Figura 3.22) e pelo RTISI (mostrado na Figura 3.23) nas situações mais críticas. Os casos em que o método MISI foi superado pelo método RTISI ocorreram principalmente quando se utiliza $N=2L$ ou $N=4L$ como número

de pontos usados nas STFTs. Para estes dois valores de N, como já comentado, os métodos RTISI e RTISI-LA apresentam melhora significativa em seus desempenhos.

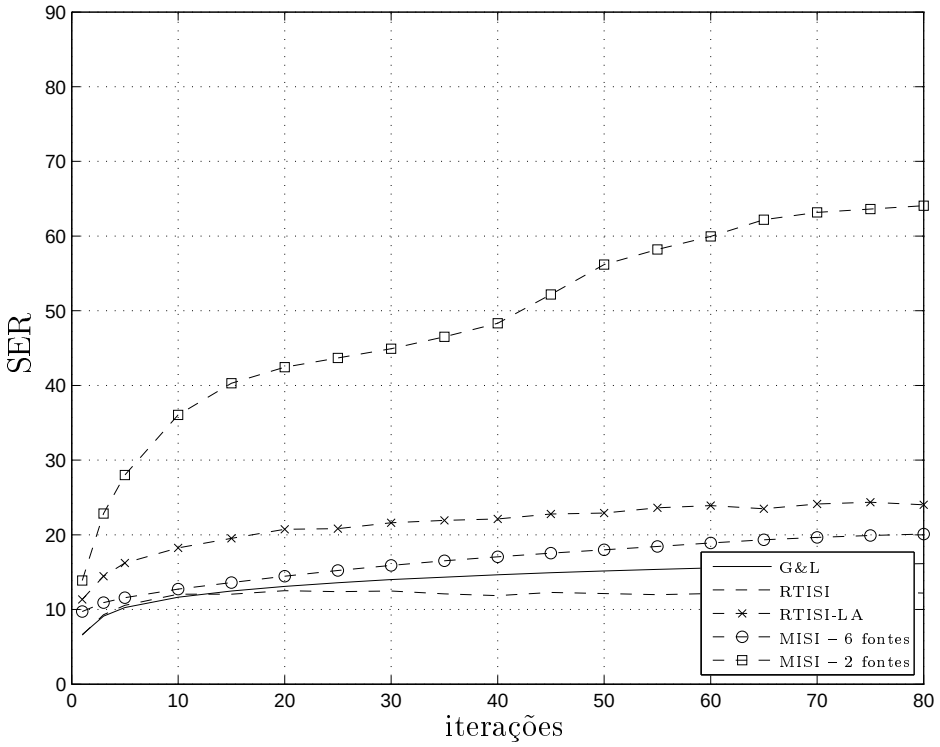

Figura 3.22: Resultados para o Sinal BR1 com DFT de 1024 pontos

No caso dos sinais de natureza percussiva, como o sinal BR6 e o sinal RN4, ou dos que possuem maior energia nas componentes graves do espectro, como os sinais BR2 e RN2, o aumento do número de fontes não fez com que o método MISI tivesse desempenho pior que os demais.

Para os casos em que uma fonte do sinal da mistura é percussiva e todas as outras são tonais, a informação de natureza percussiva é mais facilmente extraída pelo método MISI. Isto ocorre porque a informação deste tipo de sinais fica mais destacada na fase do sinal de mistura, o que pode ser comprovado ao obter um sinal temporal com base no espectro obtido ao associar a fase da mistura ao espectro de magnitude de alguma das fontes. Nesse sinal, a parcela percussiva terá sua marcação particularmente audível.

Há casos em que uma fonte possui energia concentrada na região dos graves, enquanto o resto das fontes não tem essa parte do espectro tão povoada. Nesses casos, mesmo com a adição de mais fontes, a informação de fase pertinente ao sinal mais grave pode ser facilmente reconstruída, possibilitando o desempenho superior

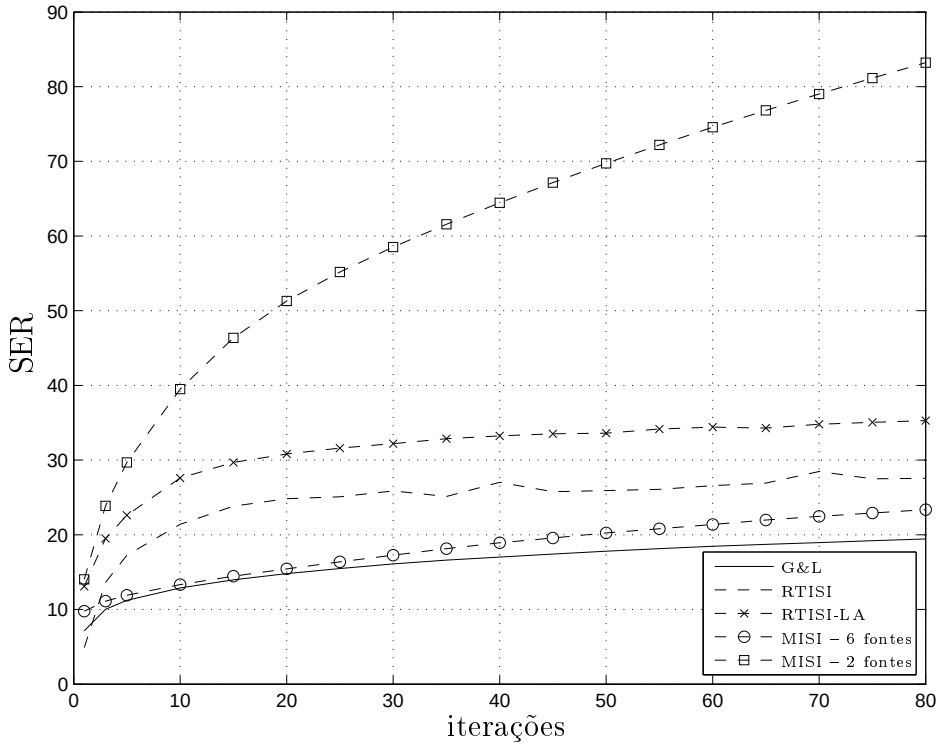

Figura 3.23: Resultados para o Sinal BR1 com DFT de 2048 pontos

do método MISI, que pode ser conferido na Figura 3.24.

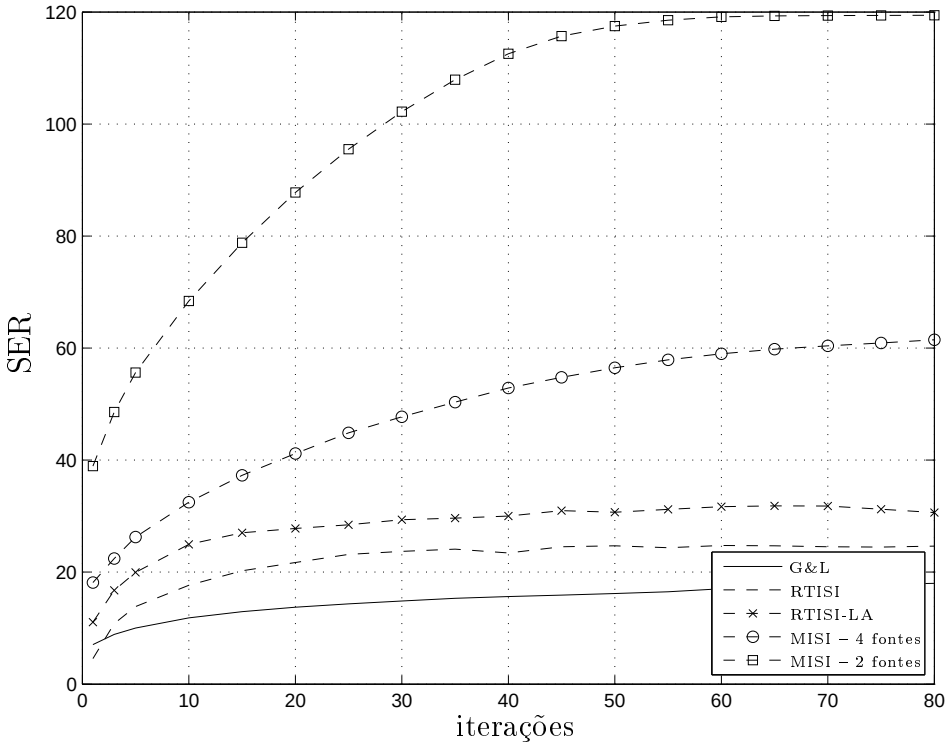

Figura 3.24: Resultados para o Sinal RN2 com DFT de 2048 pontos

3.2.4.2 Conclusão

Pôde-se concluir que o método MISI, de maneira geral, leva a resultados melhores para os três valores de N testados (1024, 2048 e 4096 pontos). No entanto, a medida que o número de fontes aumenta, o método pode atingir desempenho inferior ao RTISI-LA ou, em alguns casos, até mesmo inferior ao desempenho do método RTISI.

Para sinais percussivos ou sinais com componentes predominantes graves, o desempenho teve maior independência em relação ao número de fontes. Cabe, então, ressaltar o melhor desempenho do MISI em relação aos demais métodos em todos os casos testados com sinais que apresentam estas características.

Por fim, percebe-se que o método MISI sempre obtém resultados superiores aos algoritmo G&L, base de seu método de reconstrução. Este resultado sugere, então, que a realimentação do erro otimiza o passo de atualização da MSTFT. Assim, é possível que a inserção da realimentação do erro nas estruturas dos métodos RTISI e RTISI-LA possa levar a melhores resultados.

Capítulo 4

Conclusões e Trabalhos Futuros

Após os comentários acerca dos resultados dos testes apresentados no Capítulo 3, este capítulo visa a apresentar as conclusões gerais e possibilidades de trabalhos futuros.

4.1 Conclusões

Em todos os casos analisados, o método RTISI-LA mostrou melhores resultados na reconstrução que os métodos G&L e RTISI. Em particular, os métodos RTISI e RTISI-LA obtiveram maiores avanços em desempenho quando foram utilizados duas vezes o número de pontos da janela de análise como número de pontos das STFTs ($N = 2L$). O aumento deste número para $N = 4L$ não levou, em geral, a mais melhoras.

Para o método RTISI-LA foram testados, com os Sinais Tonais e os Sinais Percussivos, diferentes números de quadros de *Look Ahead*. Para os Sinais Tonais, verificou-se que o método é otimizado com 3 quadros de *Look Ahead*. Com relação aos Sinais Percussivos, foi verificado que o emprego de 5 quadros de *Look Ahead* é o mais adequado. Cabe-se ressaltar que, segundo indicado pelos testes, o desempenho do RTISI-LA para um mesmo número de iterações é, em geral, crescente com o número de quadros de *Look Ahead*. Porém, o aumento do número de quadros acima dos valores citados não gerou melhoras significativas de desempenho, não compensando o custo computacional.

Outro aspecto relevante observado em relação ao número de pontos N da

STFT utilizado foi quanto à medição da SER. As medições realizadas quando se utiliza $N = 2L$ (sendo, em geral, $L = 1024$) tiveram diferenças de, no máximo, 0,2 unidade com relação àquelas feitas com $N = 2^{14}$. Frente a valores típicos encontrados da SER, que variam de 7 a 100 e têm variações de até 1 unidade por conta de erros numéricos, estas diferenças tornam-se desprezíveis. No entanto, os erros numéricos não interferiram no aspecto geral dos resultados.

A utilização do critério de parada por limiar para os métodos RTISI e RTISI-LA não apresentou melhoras de desempenho para a maioria dos sinais. Com isso, não foi consolidada a ideia, discutida na Seção 2.3.2, de conferir maior eficiência aos algoritmos desses métodos.

Por fim, os testes realizados com o método MISI apresentaram, em geral, resultados melhores que os dos demais métodos. Houve, porém, casos nos quais este método foi superado pelos métodos RTISI e o RTISI-LA. Isto ocorreu ao se aumentar o número de fontes, ou seja, com o aumento do número de sinais contidos na mistura.

4.2 Trabalhos Futuros

Como citado na Seção 3.2.4.2, os testes indicaram que a realimentação do erro, presente no método MISI, otimiza o passo de atualização da MSTFT devido às melhoras com relação ao G&L, seu método-base. Desta forma, pode-se desenvolver métodos com base no RTISI e no RTISI-LA associados à atualização de erro do método MISI. Estes métodos teriam, além do desempenho superior apresentado pelo MISI, uso em tempo real.

Uma outra possibilidade está relacionada às fórmulas de atualização. Como visto, todos os métodos apresentados utilizam a abordagem de mínimos quadrados para seu desenvolvimento. Abordagens de minimização de erros com base em critérios psicoacústicos podem ser adotadas, e também utilizadas para a comparação entre resultados dos métodos.

Referências Bibliográficas

[1] TYGEL, A. F., *Método de Fatoração de Matrizes Não-Negativas para Separação de Sinais Musicais*. Dissertação de mestrado, COPPE/UFRJ, December 2009.

[2] CASEY, M., WESTNER, A., "Separation of Mixed Audio Sources by Independent Subspace Analysis". In: *Proceedings of the International Computer Music Conference*, ICMA, Berlin, Germany, August 2000.

[3] GRIFFIN, D. W., LIM, J. S., "Signal Estimation from Modified Short-Time Fourier Transform", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. ASSP-32, n. 2, pp. 236–243, April 1984.

[4] BEAUREGARD, G. T., ZHU, X., WYSE, L., "An Efficient Algorithm For Real-Time Spectrogram Inversion". In: *Proc. of the 8th Int. Conference on Digital Audio Effects (DAFX-05)*, pp. 116–121, Madrid, Spain, September 2005.

[5] GUNAWAN, D., SEN, D., "Music Source Separation Synthesis using Multiple Input Spectrogram Inversion". In: *Proc. IEEE International Workshop on Multimedia Signal Processing*, IEEE, Rio de Janeiro, Brazil, October 2009.

[6] PORTNOFF, M. R., "Time-Scale Modification of Speech Based on Short-Time Fourier Analysis", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, v. 29, n. 3, pp. 374–390, June 1981.

[7] LIM, J. S., OPPENHEIM, A. V., "Enhancement and Bandwidth Compression of Noisy Speech", *Proceedings of the IEEE*, v. 67, n. 12, pp. 1586–1604, December 1979.

[8] BEAUREGARD, G. T., ZHU, X., WYSE, L., "Real-Time Signal Estimation From Modified Short-Time Fourier Transform Magnitude Spectra", *IEEE Tran-*

sactions on Audio, Speech, and Language Processing, v. 15, n. 5, pp. 1645–1653, July 2007.