

Processamento de Segmentos Temporais

Roberto Piassi Passos Bodo

Instituto de Matemática e Estatística
Universidade de São Paulo

07 de Maio de 2012

Processamento de segmentos temporais

Apresentar diversos algoritmos no domínio do tempo que são uma combinação de blocos menores de processamento, como:

- moduladores de amplitude;
- moduladores de fase;
- filtros;
- linhas de atraso.

Esses algoritmos influenciam, principalmente, a altura e a duração do sinal de áudio de entrada.

Processamento de segmentos temporais

Alguns desses algoritmos são

- variable speed replay;
- synchronous overlap and add (SOLA);
- pitch-synchronous overlap and add (PSOLA);
- pitch shifting by time stretching and resampling;
- pitch shifting by delay-line modulation;
- pitch shifting by PSOLA and formant preservation;

Variable speed replay



Variable speed replay

Reprodução com velocidade mais rápida:

- diminui a duração do som;
- aumenta a altura do som.

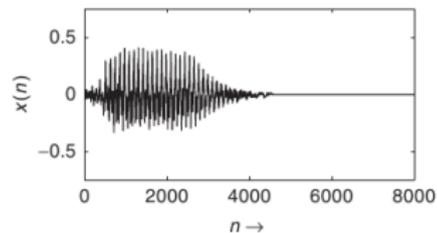
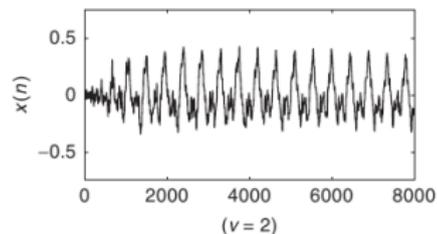
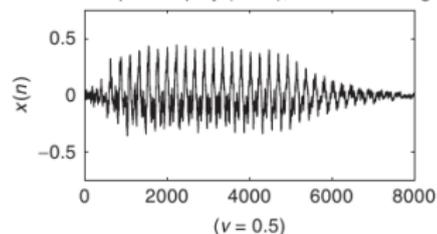
Reprodução com velocidade mais lenta:

- aumenta a duração do som;
- diminui a altura do som.

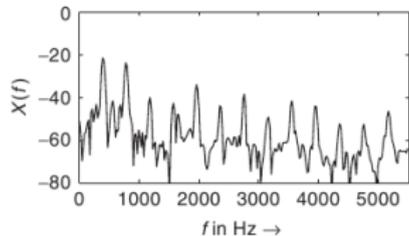
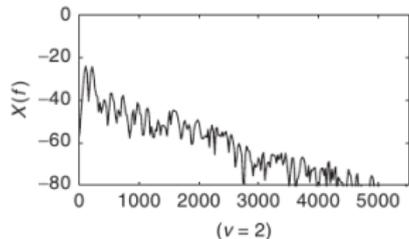
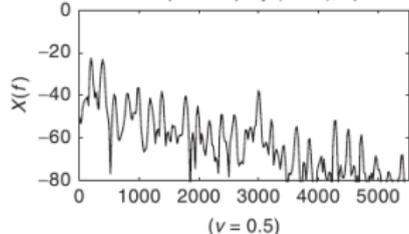
Compressão/expansão no domínio do tempo nos leva a compressão/expansão no domínio das frequências.

Variable speed replay

Variable speed replay ($v = 1$), time domain signals



Variable speed replay ($v = 1$), spectra



Variable speed replay

Período de tempo original: $n * T_{s,in}$

Período de tempo da reprodução: $n * T_{s,replay} = n * T_{s,in} / v$

onde v é a velocidade relativa,

$T_{s,in}$ é o período de amostragem inicial e

$T_{s,replay}$ é o período de amostragem da reprodução.

Variable speed replay

Método de implementação:

modificar a frequência de amostragem do sinal de saída $f_{s, replay} = f_{s, in} * v$.

Se a saída for analógica: podemos modificar a frequência de amostragem do conversor DA.

Se a saída for digital: podemos fazer uma conversão entre $f_{s, replay}$ e $f_{s, out}(= f_{s, in})$.

Variable speed replay

Se a saída for digital:

Expansão do tempo

$$\Rightarrow v < 1$$

$$\Rightarrow f_{s,in} > f_{s,replay} < f_{s,out}$$

\Rightarrow devemos fazer uma superamostragem do sinal

Compressão do tempo

$$\Rightarrow v > 1$$

$$\Rightarrow f_{s,in} < f_{s,replay} > f_{s,out}$$

\Rightarrow devemos fazer uma decimação do sinal

Time stretching

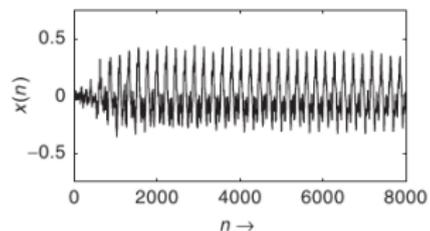
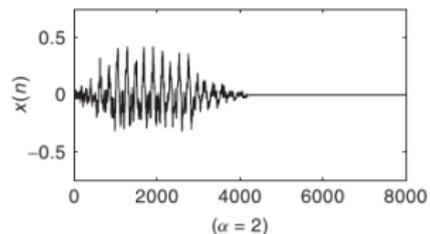
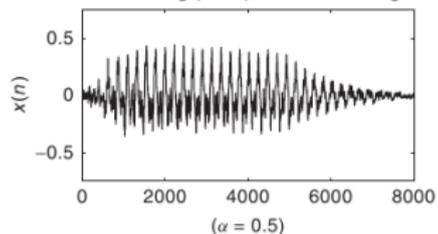
O método anterior altera a duração do sinal de entrada, mas o transpõe simultaneamente.

Queremos fazer uma mudança na escala temporal sem mudar o espectro do sinal.

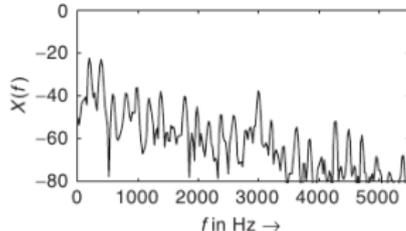
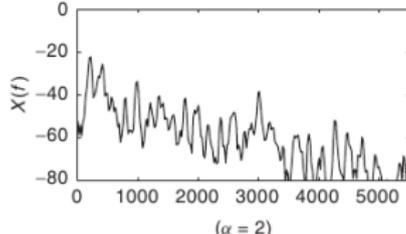
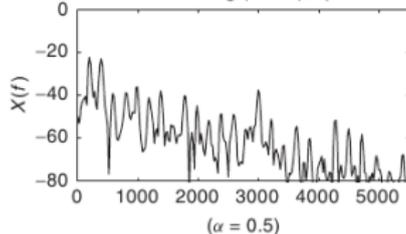
Basicamente, um sinal de M amostras deve ter um novo comprimento $M' = \alpha * M$, onde α é o fator de escala.

Time stretching

Time stretching ($\alpha = 1$), time domain signals



Time stretching ($\alpha = 1$), spectra



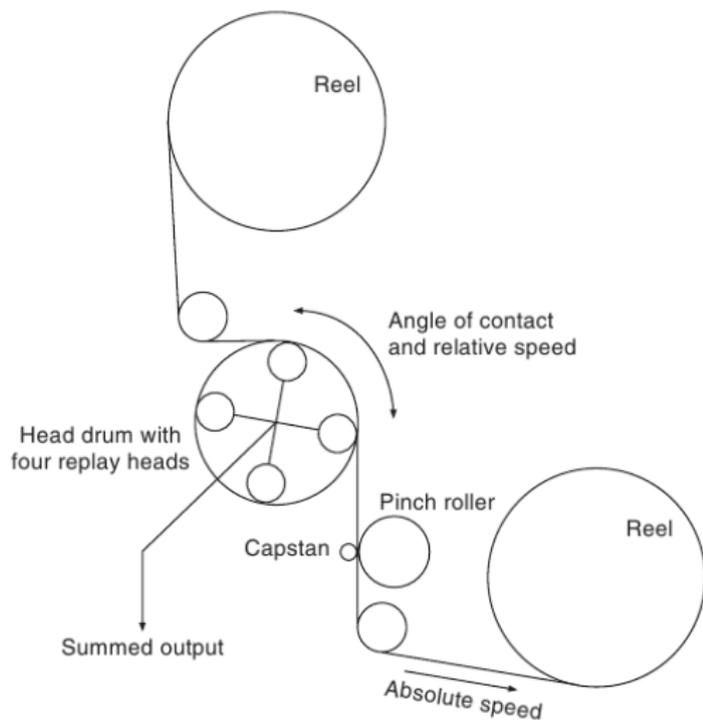
Time stretching

Aqui a escala de tempo não corresponde à escala matemática e, sim, dos atributos de tempo percebidos.

Queremos a mesma sequência de eventos acústicos sendo reproduzidos em um novo padrão de tempo.

No caso de um discurso humano, queremos escalar o ritmo da fala e, não, as frequências.

Time stretching



Time stretching

Ideia básica de implementação:

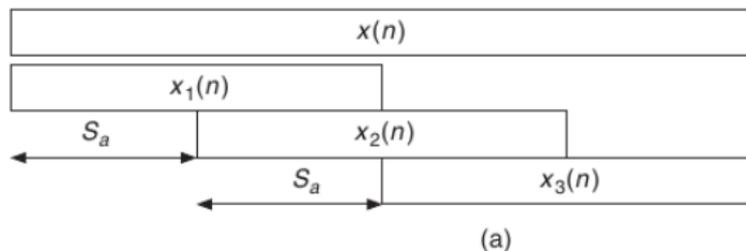
- dividir o sinal de entrada em segmentos no domínio do tempo;
- repetir alguns segmentos se o som deve ser alongado;
- descartar alguns segmentos se o som deve ser encurtado.

Problemas:

- descontinuidades de amplitude: evitadas com uma sobreposição parcial dos blocos.
- descontinuidades de fase: evitadas com um alinhamento apropriado dos blocos.

Synchronous OverLap and Add (SOLA)

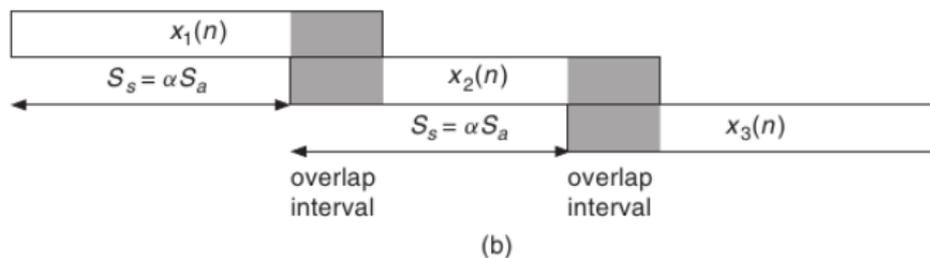
(1) Segmentar o sinal de entrada em blocos sobrepostos de tamanho N deslocados no tempo por S_a amostras.



Synchronous OverLap and Add (SOLA)

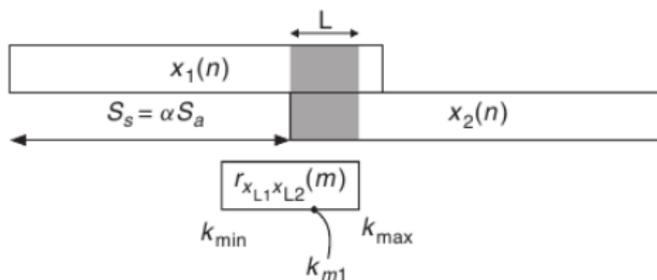
(2) Reposicionar os blocos com um novo deslocamento no tempo

$S_s = \alpha * S_a$ usando o fator de escala α .



Synchronous OverLap and Add (SOLA)

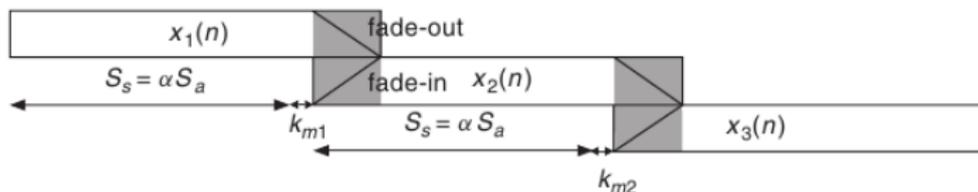
(3) Calcular a correlação entre os segmentos no intervalo sobreposto de comprimento L para obter o índice k_m para o qual a correlação possui o valor máximo.



(c)

Synchronous OverLap and Add (SOLA)

(4) Utilizar k_m para sincrozinar os blocos, aplicar *fade-out* em um, *fade-in* em outro e somá-los formando o sinal de saída.



(d)

Pitch-Synchronous OverLap and Add (PSOLA)

Baseado no SOLA.

Hipótese: o sinal de entrada é caracterizado por frequências únicas a cada instante de tempo, como na voz humana ou em instrumentos monofônicos.

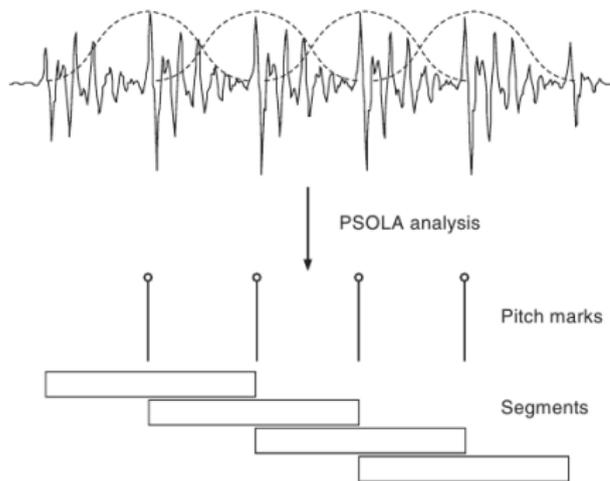
O PSOLA explora o conhecimento das alturas para sincronizar os segmentos temporais.

O algoritmo é dividido em duas fases: uma de análise e segmentação do sinal e uma de síntese da versão escalada no tempo.

Pitch-Synchronous OverLap and Add (PSOLA)

Algoritmo de Análise:

- (1) determinar os *pitch periods* $P(t)$ e os *pitch marks* t_i do sinal de entrada;
- (2) extrair um segmento centralizado em cada *pitch mark* t_i usando uma janela de Hanning de comprimento $L = 2 * P(t_i)$.



Pitch-Synchronous OverLap and Add (PSOLA)

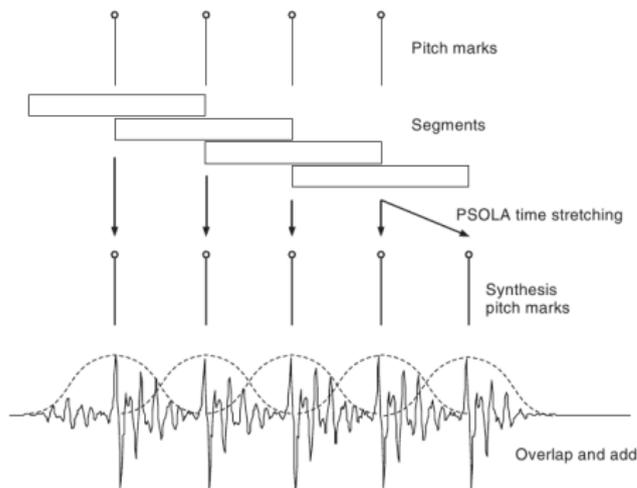
Algoritmo de Síntese: para cada *pitch mark* \tilde{t}_k da saída

(1) escolher o segmento i correspondente da entrada minimizando

$$|\alpha t_i - \tilde{t}_k|;$$

(2) sobrepor e somar o segmento selecionado;

(3) determinar o instante de tempo $\tilde{t}_{k+1} = \tilde{t}_k + \tilde{P}(\tilde{t}_k) = \tilde{t}_k + P(t_i)$ onde o próximo instante de tempo será centralizado.



Pitch marks

A determinação dos *pitch periods* e dos *pitch marks* não é um problema trivial.

O posicionamento dos *pitch marks*, calculados a partir dos *pitch periods*, determina os centros das janelas de segmentação do PSOLA e, com isso, a qualidade do som do algoritmo.

Pitch marks

Determinação dos *pitch marks*:

- localizar o pico global do sinal de entrada e defini-lo como o primeiro *pitch mark* t_i ;
- calcular os *pitch marks* à direita de t_i usando os *pitch periods* como uma primeira estimativa;
- fazer um refinamento dessa localização procurando o máximo local dentro de uma determinada região:
$$t_j = \max([t_i + \delta * P_o, t_i + (2 - \delta) * P_o, \text{ onde } P_o \text{ é o período e } \delta \text{ é um fator entre } 0,5 \text{ e } 0,9.$$
- fazer o mesmo processo até a extremidade direita do frame e o processo análogo até a extremidade esquerda do frame.

Pitch shifting

Transposição: repetir uma melodia após deslocar a altura dela por um intervalo fixo.

Timbre: "característica sonora que nos permite distinguir se sons de mesma frequência foram produzidos por fontes sonoras conhecidas e que nos permite diferenciá-las." [Wikipedia]

"O timbre de um som depende profundamente da organização de seu espectro." [DAFX]

Podemos escolher entre "transposição com timbre variável" e "transposição com timbre constante".

Pitch shifting

O que é pitch shifting?

Modelo: estudo do canto humano.

A altura é determinada pelas cordas vocais. O timbre, pelas cavidades vocais.

Excitação (relacionada com as frequências do espectro) é modificada por uma ressonância (relacionada com os formantes).

Um cantor tem a possibilidade de modificar a altura e os formantes independentemente.

Em uma implementação mais meticulosa devemos considerar ambos os aspectos.

Harmonizer

Um dos primeiros dispositivos digitais a fazer sucesso. Fabricado pela Eventide nos anos 70.

Era capaz de modificar a altura mantendo a duração original. Possuía uma razão de transposição maior do que dispositivos analógicos.

Principal limitação: qualidade característica para os sons processados (percebida tanto no domínio do tempo, quanto no das frequências).

Implementações digitais: quanto maior a razão de transposição, mais perceptíveis eram as modificações no timbre.

Harmonizer

Modulação do sinal por janelas:

O espectro do sinal de entrada é convoluido com o da janela.

Comprimento da janela: quanto mais longo, menos perturbador à modulação.

Forma da janela: trapezoidal terá um efeito mais forte do que uma janela mais suave (por exemplo, Hanning).

No entanto, janelas mais longas acabam gerando cópias audíveis do sinal de entrada e janelas mais suaves não reproduzem bem os transientes.

Pitch shifting by time stretching and resampling

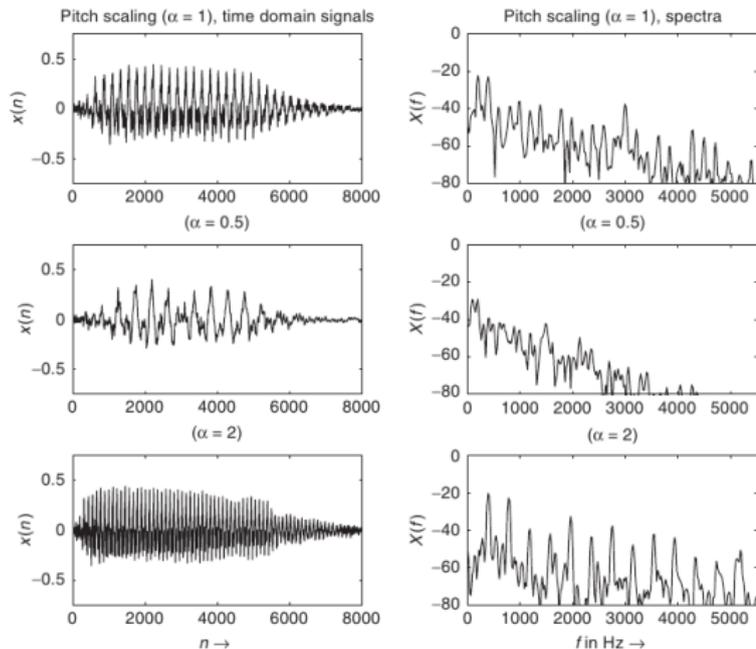
A reprodução com velocidade variável faz uma reamostragem do sinal no domínio do tempo.

O espectro do som é comprimido e expandido no domínio das frequências.

As relações harmônicas $f_i = i * f_0$ continuam as mesmas, mas são escaladas conforme $f'_i = \alpha * f_i$. As amplitudes dos harmônicos continuam as mesmas $a'_i = a_i$.

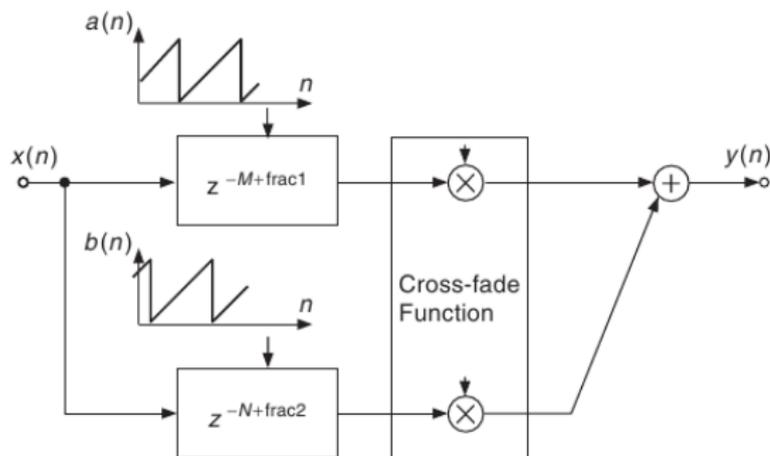
Podemos, então, aplicar um algoritmo de mudança de escala temporal.

Pitch shifting by time stretching and resampling



Pitch shifting by delay-line modulation

Método: sobreposição e soma com duas linhas de atraso com tempo variável.



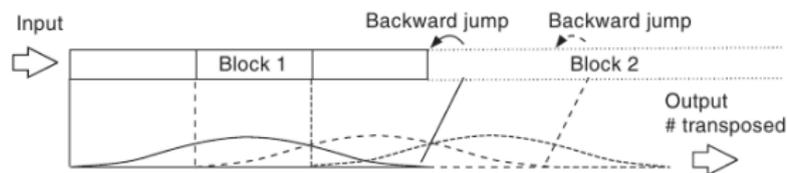
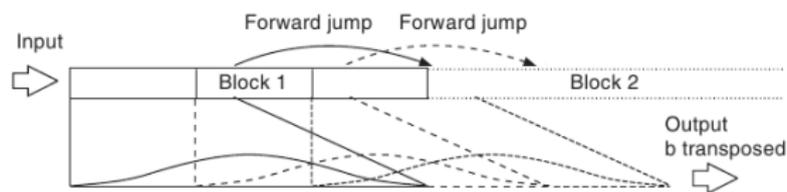
Pitch shifting by delay-line modulation

- O sinal é dividido em blocos pequenos que são lidos com velocidades diferentes;
- Dois blocos são lidos simultaneamente com um atraso igual à metade do comprimento do bloco;
- É feito um *cross-fade* de um bloco para outro ao final de cada bloco;
- O tamanho das linhas de atraso é modulado por uma função dente de serra.

Pitch shifting by delay-line modulation

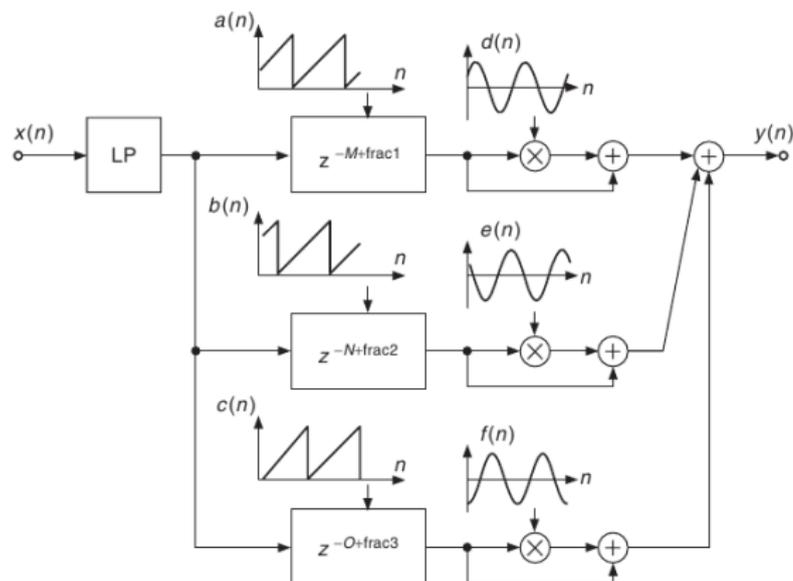
Aprimoramento do método:

(1) sinal de entrada é dividido em blocos que são reamostrados, modulados em amplitude e somados, produzindo um sinal de mesmo tamanho do sinal de entrada.



Pitch shifting by delay-line modulation

(2) os sinais de modulação formam um sistema de três funções deslocadas de 120° . A soma dessas funções sempre é constante.



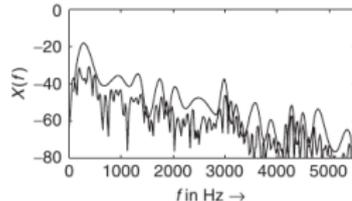
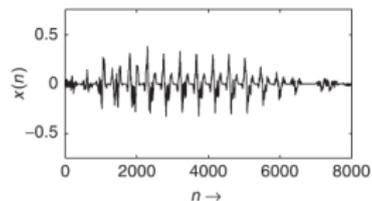
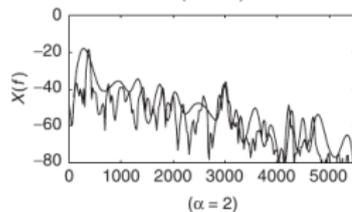
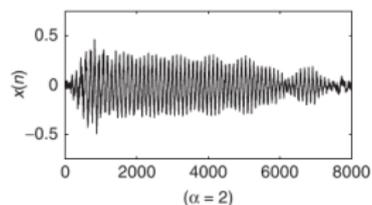
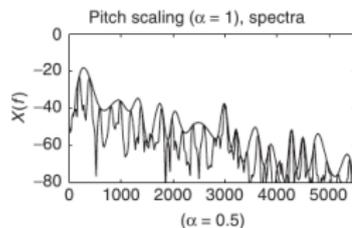
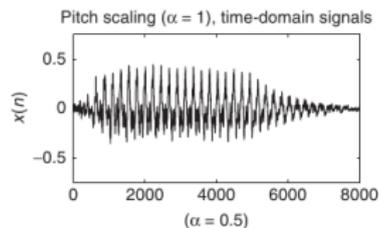
Pitch shifting by PSOLA and formant preservation

Características:

- Reamostragem no envelope do espectro;
- os harmônicos são escalados $f'_i = \beta * f_i$;
- as amplitudes são escaladas também $a'_i = env(f'_i) \neq a_i$;

PSOLA é convenientemente utilizado para deslocar a altura de uma voz sem alterar a altura dos formantes.

Pitch shifting by PSOLA and formant preservation



Pitch shifting by PSOLA and formant preservation

Ideia básica do algoritmo é mudar a escala temporal das posições dos *pitch marks* sem modificar a forma de onda do segmento.

Podemos pensar no modelo de produção da fala como um trem de pulsos filtrados por um filtro correspondente ao trato vocal.

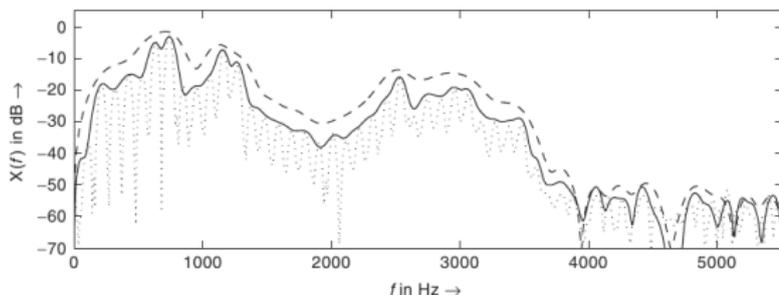
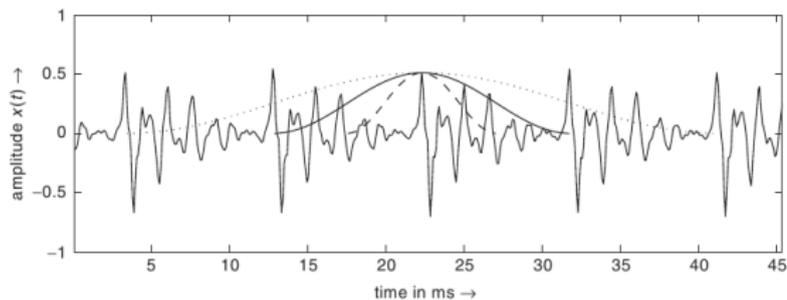
O segmento de entrada corresponde à resposta impulsiva desse filtro e determina a posição dos formantes (que não deve ser modificada).

A distância dos *pitch marks* determina o período do discurso (que deve ser modificado).

Pitch shifting by PSOLA and formant preservation

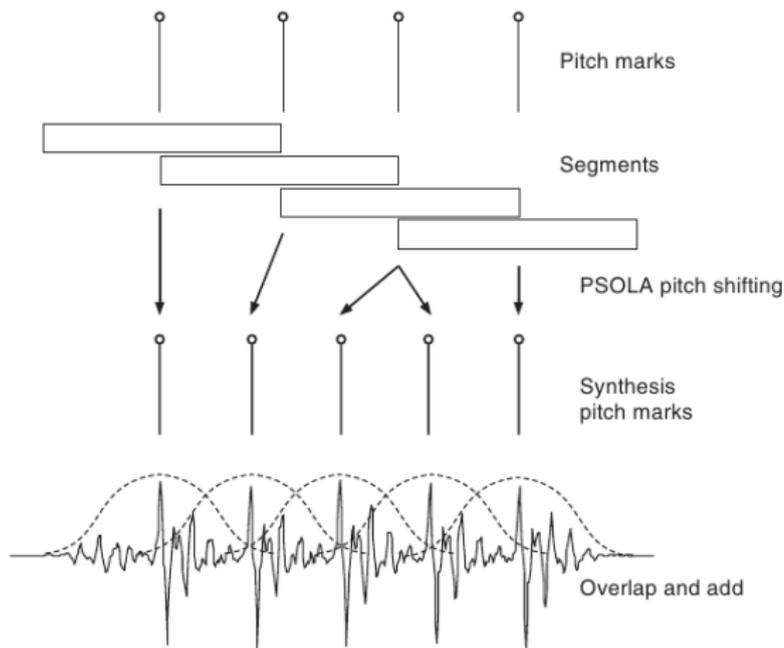
A análise do PSOLA extrai a resposta impulsiva do filtro.

Janelas mais longas/espectro mais fino. Janelas mais curtas/formantes mais "borrados".



Pitch shifting by PSOLA and formant preservation

Síntese do PSOLA:



That's all folks!

Obrigado! =)