The Short-time Fourier Spectrump

- Fourier transform이란 무엇인가

- 일종의 수학적 단계
- 'frequency domain의 함수', '하나의 함수를 다른 함수로 변환하는 과정'이라는 두 의미 지님
- **연속 시간의 아날로그 파형을 infinite Fourier series(무한 푸리에 급수)의 합으로** 만드는 것
- infinite Fourier series는 특정 amp와 phase를 가지는 사인파로 만들어짐
- **입력 신호**를 대응하는 **스펙트럼으로 전환하는 것**
- **시간영역의 함수를 주파수영역의 함수로 변환하는 것**
- short time Fourier transform(STFT): FT를 실제 녹음된(유한의) 소리에 적용하기 위해 만든 것

Windowing the Input Signal

- STFT는 입력 신호에 일련의 time window를 부과하는 것
- window는 스펙트럼 분석을 위한 것으로 특별한 형태의 envelope을 의미
  window의 duration은 1ms에서 1second
- 이 조각들의 스펙트럼을 각각 분석함으로서 실시간 변화하는 스펙트럼의 연속 값 획득
- windowing 과정을 통해 스펙트럼 측정값이 왜곡되는 효과 발생
 : 입력신호가 순수하게 분석되지 않고, 입력 신호와 window의 산물인 입력 신호와 윈도우 신호 스펙트럼의 convolution이 나타남

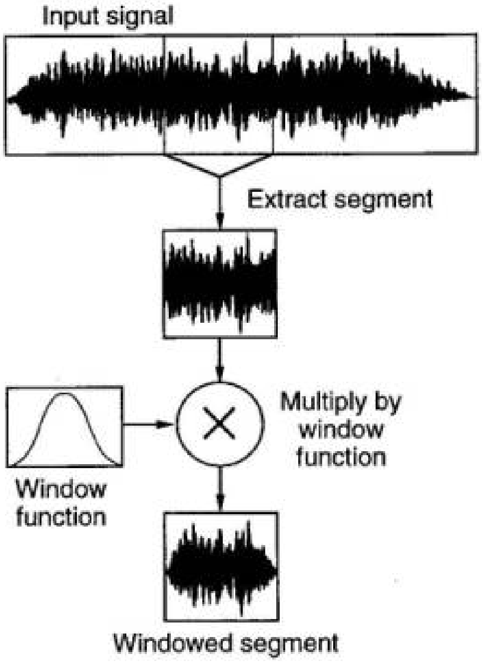


Figure 13.7 Windowing an input signal

Operation of the STFT

- Windowing 이후 STFT의 각 조각에 discrete Fourier transform(DFT) 적용
- DFT(discrete-frequency spectrum)는 일종의 Fourier transform 알고리즘
  이것을 이용해 discrete 혹은 샘플 된 사운드 취급
- DFT의 출력 신호는 discrete-frequency spectrum 즉 동일한 간격에서 주파수들이 갖는 에너지 값
- Fast Fourier Transform(FFT)은 DFT의 단순한 방법[1]
  따라서 STFT의 가장 실질적 이행은 FFT알고리즘을 각 (window 된) 조각에 적용하는 것

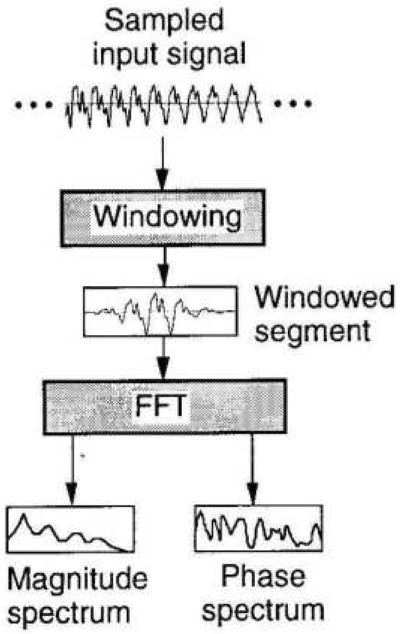


Figure 13.8 Overview of the short-time Fourier transform(STFT)

- FFT에 의해 만들어진 각 block의 데이터를 프레임이라 함(film의 연속된 프레임과 유사)
- 프레임의 두 가지 요소
① Magnitude spectrum: 모든 분석된 주파수 요소의 amp 값
② Phase spectrum: 모든 분석된 주파수 요소의 초기 phase 값
- histogram[2]으로 횡좌표에서 각 주파수 요소를 위한 수직 정보를 이용해 스펙트럼 인식 가능
- 수직선이 의미하는 것은 magnitude spectrum의 경우 amp 값
  phase spectrum의 경우 starting phase(-π에서 π)
- phase spectrum이 -π에서 π범위에서 normalize 된 경우 wrapped phase라 함
- 많은 신호에서 random function으로 보임. unwrapped phase projection이 시각적으로 더욱 의미
- 입력 신호에 STFT의 적용은 시간에 따라 변하는 스펙트럼의 연속 프레임 생성

1) FFT는 DFT를 빠르게 구현하기 위한 알고리즘을 총칭하는 것
2) 도수분포표에 따라 계급을 밑면, 도수를 높이로 하는 직사각형의 넓이로 나타낸 그래프

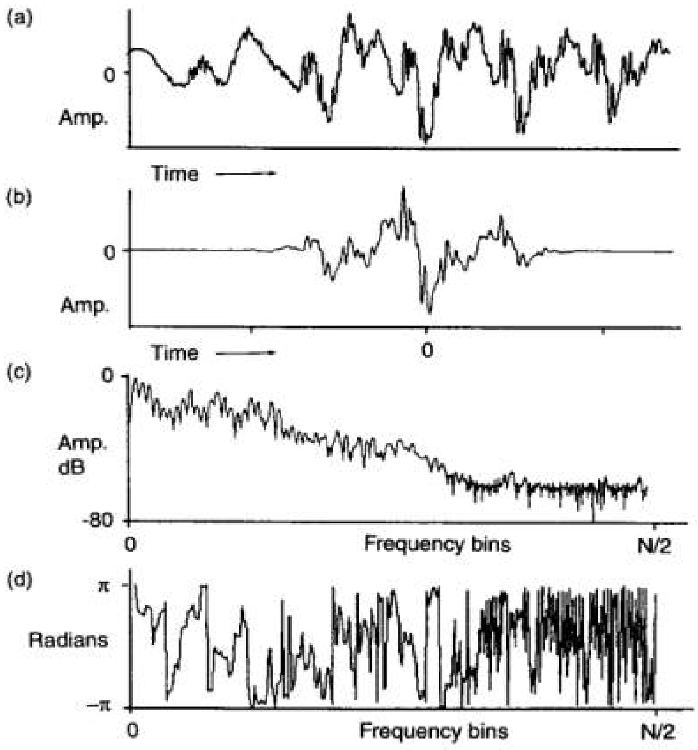


Figure 13.9 STFT signals (a) Input waveform. (b)Windowed segment. (c) Magnitude spectrum. (d) Phase spectrum.

Overlap-add Resynthesis from Analysis Data

- time domain의 신호를 재합성하는 방법

: STFT는 각 프레임에 IDFT(inverse discrete Fourier transform)를 적용하여 스펙트럼 요소로부터 윈도우된 파형의 단편 재구축

① IDFT는 analysis window로서 각각의 magnitude와 phase의 envelope을 가지는 time domain 신호 생성

② window를 겹쳐 쌓거나 더함으로서 원 소리와 유사한 신호 생성

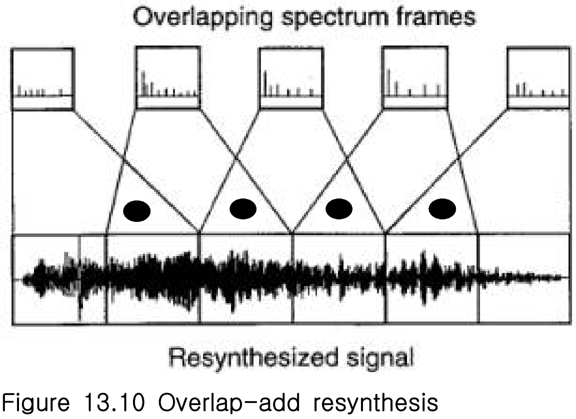
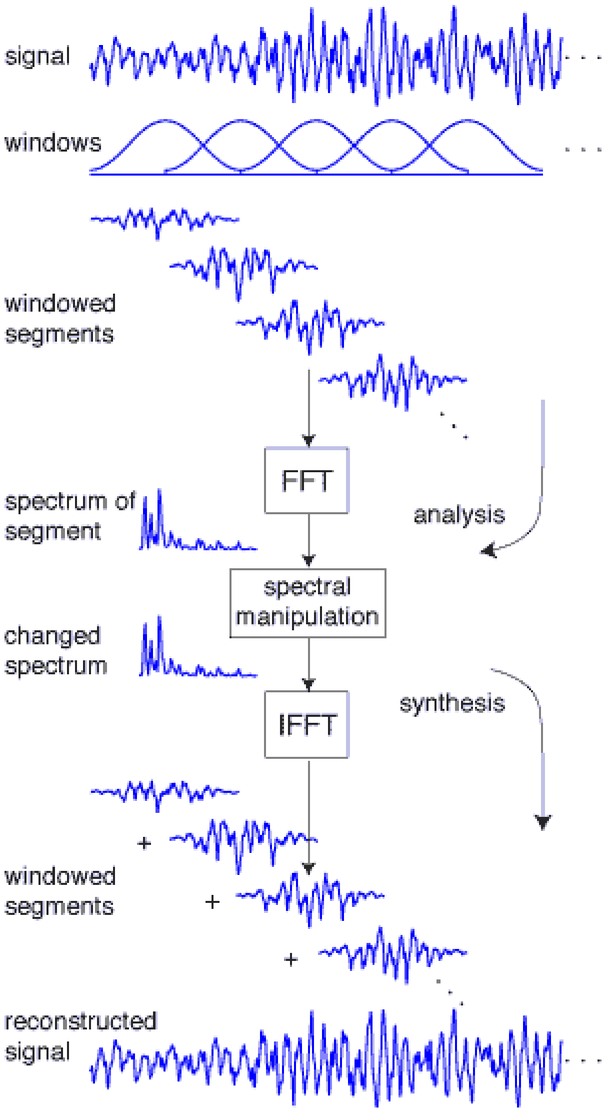


Figure 13.10 Overlap-add resynthesis

● = overlapping spectrum frame

- 이론적으로 STFT의 재합성은 identity operation(일치연산)[3].
- 이것이 실제 identity operation이라면 STFT를 이용해 어떤 손실 없이 신호 copy 가능
- 하지만 좋은 STFT의 경우도 적은 양의 정보를 손실. 이 손실은 STFT를 통과한 후 들을 수 없는 정도



STFT http://eceserv0.ece.wisc.edu/~sethares/htmlRT/vocoders/phasevocoder.html

Limits of Overlap-add Resynthesis

- overlap-add(OA)를 이용한 재합성은 (변화 많은)음악에 적용하는데 제약 존재
- 한계 발생 이유: OA진행의 경우 윈도우의 합이 완벽하게 지속적인 경우를 위해 고안
- Allen과 Rabiner(1977): OA의 방법은 들을 수 있는 부수적 효과를 만듦
- 이러한 원치 않은 인공적임을 줄이는 한 가지 방법은 분석 단계에 연속된 window 사이 많은 overlap을 사용하는 것(다음 section에서 설명)
- improved overlap-add 재합성 방법은 이러한 문제를 해결하는 또 다른 방법

3) identity operation: In mathematics, an identity function is a function that always returns the same value that was used as its argument. In terms of equations, the function is given by f(x) = x.

Why Overlapping Windows[4]?

- STFT에서 overlapping 분석의 유도는 혼란을 야기할 가능성을 지니지만 이론적으로 어떤 길이의 조각이든 분석 가능하며, 정확하게 분석 데이터를 이용해 그 조각을 재합성하는 것 역시 가능
  예) 30분 길이의 윈도우를 사용하여 Stravinsky의 Le sacre du printemps를 분석하고 이 분석에서 전체 작품을 다시 합성 가능

** short time overlapping을 하는 이유
  ① windowing의 첫 번째 역할은 스펙트럼의 visualization
  : 30분 분량의 44.1KHz 모노 사운드를 분석할 경우 79만개 이상의 점으로 이루어진 스펙트럼 생성. 이 많은 양의 스펙트럼의 시각적 검사는 30분 동안 발생하는 모든 frequency 제시. 하지만 언제 정확하게 발생하는지 알려주지 못함
  → 이러한 일시적 정보는 magnitude와 phase spectra의 수학적 결합과 관련된 깊이 있는 정보 알려주지 못함
  조각을 나누어 분석을 제한함으로서 적은 양의 점으로 분석
  ② 메모리의 보존
  : 30분 분량의 16비트 샘플의 경우 컴퓨터가 FFT를 하는 동안 input을 받아들이기 위해 적어도 79만 16비트 단어를 위한 RAM의 컴퓨터가 필요. input bite size의 조각으로 나눔으로서 한꺼번에 여러 작은 조각을 FFT하는 것 쉬워짐
  ③ 빠른 결과를 획득
  : Stravinsky 작품에서 입력 신호를 읽기 위해 약 30분 소요(+ FFT 분석 시간).
  input을 windowing할 경우 입력 신호가 읽혀지고 몇 초 후에 초기 결과 획득 가능하며 실시간 스펙트럼 분석을 위해 적용

** 왜 window를 겹치는 것인가?
  : 부드러운 bell 모양의 윈도우는 windowing에서 발생하는 distortion을 최소화 함. 물론 bell 모양의 window도 결함 없는 신호를 얻기 위해 다소 오버랩 되어야 함. 오버랩 요소의 증가는 스펙트럼의 oversampling에도 적용되기 때문에 time-stretching이나 cross-synthesis 같은 변형에서 발생할 수 있는 aliasing 방지
  - 입력 신호를 변형하고자 하는 경우 8개 이상의 오버랩 요소 필요
  - 윈도우의 setting과 길이에 관해서는 이후에 논함

Oscillator Bank Resynthesis

- Sinusoidal additive resynthesis(SAR)(oscillator bank resynthesis)는 overlap-add(중첩-연결)와 다름
- OA 재합성 모델은 각 프레임에서 사인파를 더하는 것이고
  SAR는 프레임 경계를 가로지르는 음량과 주파수 envelope에 의한 oscillator bank를 적용하는 것

4) When the length of a data set to be transformed is larger than necessary to provide the desired frequency resolution, a common practice is to subdivide it into smaller sets and window them individually. To mitigate the "loss" at the edges of the window, the individual sets may overlap in time.

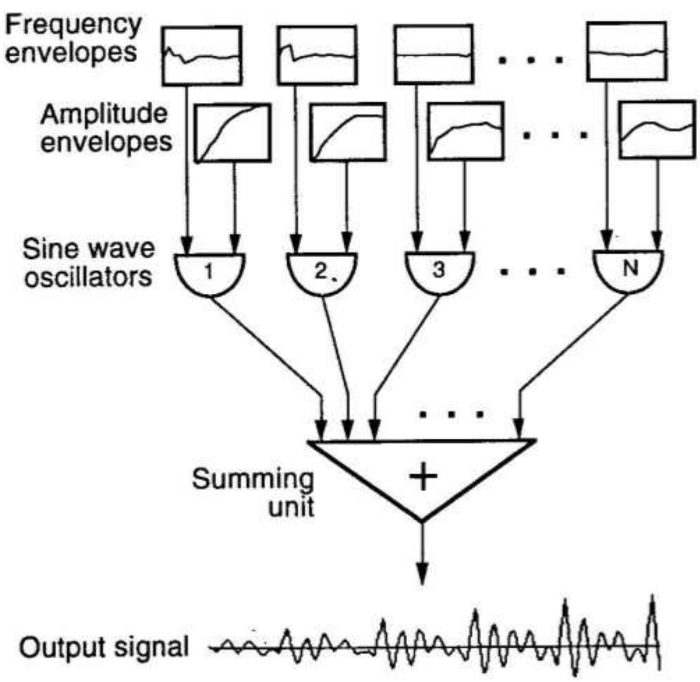


Figure 13.11 Oscillator bang synthesis

- 이것은 분석 데이터가 이런 envelope보다 먼저 변환되어야한다는 것을 암시
- 다행히 분석 데이터(크기와 위상)에서 합성 데이터(음량과 주파수)로의 변환을 위한 소요 시간은 크지 않음
- SAR 모델의 이점: 음악적 변형에 더 강함
- SAR 모델의 단점: 계산적으로 비효율적
- phase vocoder tracking 역시 가산 사인파합성을 위해 주파수의 envelope을 구성한다는 점에서 SAR 방법처럼 보일 수 있음(phase vocoder 섹션에서 좀 더 세밀하게 접근)

Analysis Frequencies

- STFT은 window된 입력 신호에 filter bank[5)]를 동일한 frequency 간격으로 적용한 것
- 주파수들은 sampling frequency/N의 정수배로(i.e, harmonics) 나뉘며, 여기서 **N이 분석된 부분의 크기**

예) 샘플링 주파수가 50KHz이고 윈도우 길이가 1000 sample이라면, 분석 주파수들은 0Hz에서 시작하여, 50,000/1000 = 50Hz 간격으로 위치. 0Hz에서 신호의 direct current나 DC offset을 측정하고, **음량 0을 중앙으로 위나 아래로 전체 신호를 바꿀 수 있다.**

- 오디오 신호는 sampling rate의 반으로 한계가 정해지기 때문에 - 예의 경우 25KHz, 우리는 분석 bins의 단지 반을 인식
  (bin은 시그널 프로세싱의 용어로 주파수 채널)

5) A filter bank is an array of band-pass filters that separates the input signal into several components

Time/Frequency Uncertainty

- 모든 window된 스펙트럼 분석 시 시간과 주파수 분석에 기본적인 불확실성(Uncertainty)원리 존재
- time domain에서 더 정확한 분석을 원할 경우 주파수 분석을 희생해야 함
→ 하나의 이벤트가 정확한 시간에서 발생하지만 우리는 정확하게 그 시간의 주파수를 말할 수 없다
- 반대로 frequency domain에서 정확한 분석을 원할 경우 시간 분석 희생해야 함
→ 긴 시간 간격에서 만 주파수 정확하게 측정 가능
- Fourier 분석의 결과를 해석하는데 이러한 관계 파악이 중요

Periodicity Implies Infinitude

- Fourier 분석은 하나의 신호가 하나의 주파수를 갖고, 무한한 duration이어야 한다는 추상적 전제에서 출발
- Gabor: frequency 개념이 무한 신호만을 위한 개념이라면 변화하는 주파수의 개념은 불가능
- 하지만 사고실험(thought experiment)을 통해 추상적인 Fourier의 개념 중 하나의 특징이해 가능 (Figure 13.12)
: sound editor를 사용해 디지털 시스템이 가진 time domain의 zoom in/out
 ① zoom in: 가장 짧은 순간에 각각의 sample point 볼 수 있음(13-12a). 이 샘플이 발생한 시간을 정확히 알 수 있으며 정밀한 결과 인식 가능. 하지만 어떤 파형의 부분인지 알 수 없음
 ② zoom out: 더 많은 분석 샘플 볼 수 있음. 그 주파수가 어느 부분인지 인식 쉬움
- Fourier 분석은 모든 구간의 스펙트럼을 동시에 계산하기 때문에 긴 구간을 보여주는 스펙트럼의 경우 특정 주파수가 생긴 시간을 정확히 알기 힘듦
- Frequency를 정밀하게 분석할 경우 시간적으로 부정확해짐

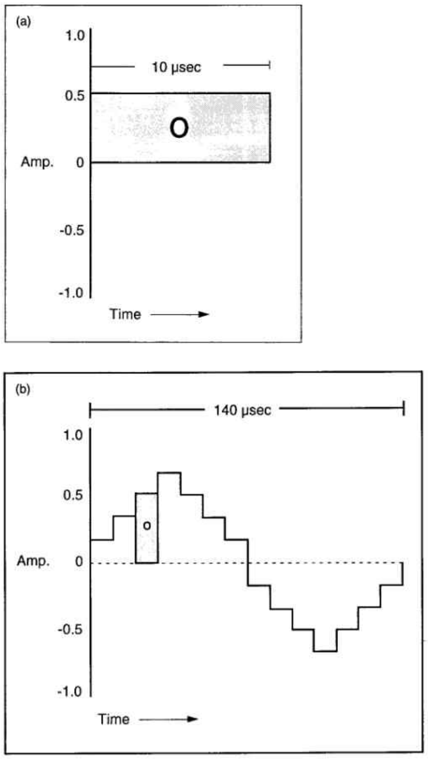
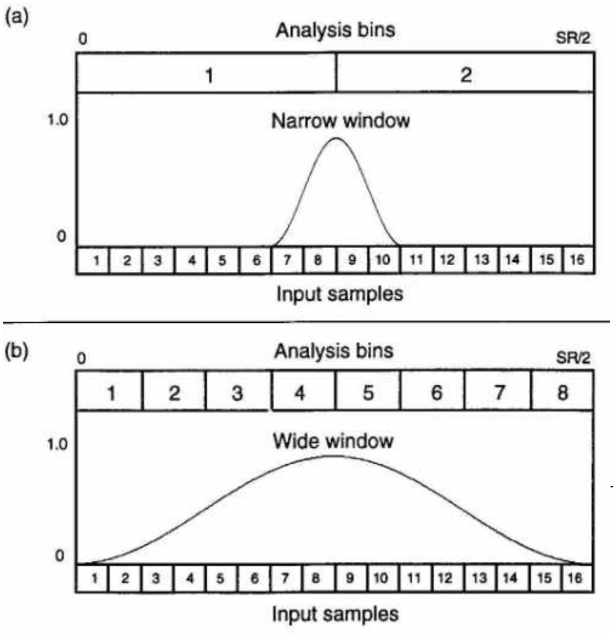


Figure 13.12 Frequency uncertainty at a small timescale

Time/Frequency Tradeoff

- FFT는 가청 주파수를 *N/2* frequency bins로 분할
- N은 분석 윈도우의 샘플 길이. ∴ frequency bins의 개수와 분석 윈도우의 길이 간 관계 존재
예) N=512 sample의 경우 분석될 수 있는 주파수 개수는 256
→ 44.1kHz 경우: 0Hz~22.05kHz(Nyquist frequency)의 대역폭을 균등하게 나눈 256 bins 획득



→ 4개 샘플의 좁은 윈도우는 2개의 주파수 분석 (2개의 bin)

→ 16개 샘플의 넓은 윈도우는 8개의 주파수 분석 (8개의 bin)

Figure 13.13 Relationship of window size to the number of frequency analysis bins

Table 13.1 Time vs frequency resolution in windowed spectrum analysis

Length of time window(in ms)	Frequency resolution (analysis band width)(in Hz)
1	1000
2	500
3	330
10	100
20	50
30	33
100	10
200	5
300	3
1000 (1sec)	1
2000	0.5
3000	0.3

- 시간과 주파수 resolution의 균형에 관한 설명
① high time accuracy: 22 frequency bin에 만족해야 함
오디오 bandwidth를 22개의 bin으로 0부터 22.06kHz까지 나누면 22,050/22, 약 1000Hz의

frequency resolution 획득. 즉 1ms 단위로 이벤트가 발생한 시간을 알고자 한다면 frequency resolution은 1000Hz

② high resolution in frequency: 분석 간격을 30ms으로 늘릴 경우 33Hz bandwidth 내에서 frequency 찾을 수 있음

- 이러한 windowed STFT analysis의 제한 때문에(양쪽 면에서 모두 높은 분해도로 분석하기 위해) time-domain과 frequency domain analysis를 혼합한 *multiresolution analysis*, non-Fourier 방법 사용(later section)

Frequencies in between Analysis Bins

- STFT는 audio bandwidth를 중심으로 동일 간격으로 위치하고 있는 주파수의 불연속적 세트에 대해서만 알 수 있음
- 이 주파수의 간격은 분석 window의 길이에 따라 결정되며, 이 길이는 기본 주기
- harmonic 혹은 quasi-harmonic sound의 경우 잘 적용 됨
- 하지만 동일한 간격에서 이루어진 STFT의 analysis bins 사이 빠진 주파수의 경우, 문제 발생
  (gong이나 snare drum처럼 noisy한 sound를 갖는 inharmonic sound의 경우)
- f를 분석한다고 가정

① f가 analysis channel의 중앙과 일치할 때, 모든 에너지는 이 채널에 집중하여 정확하게 측정 가능

② f가 중앙에 가깝기는 하나 정확히 일치하지 않을 때, 에너지는 다른 모든 analysis channel로 흩어지나, f에 가깝게 집중되는 현상 존재

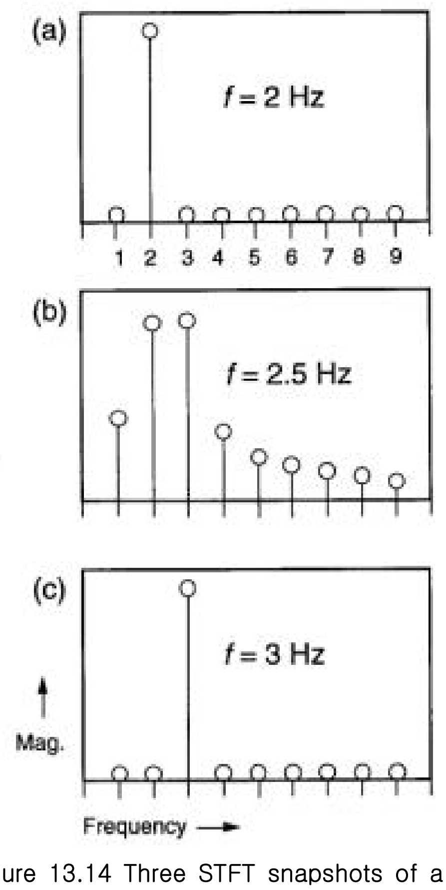


Figure 13.14 Three STFT snapshots of a sound changing frequency from 2 to 3 Hz

- STFT는 1 Hz 간격의 analysis bins
- input 주파수가 2.5Hz일 때, 동일하게 떨어진 frequency bins 으로 나타나며 에너지는 스펙트럼 전체로 분산
- 이것은 스펙트럼 측정에서 STFT에 의해 만들어진 신뢰할 수 없는 소스로 알려짐
- bin사이에 하나 이상의 요소가 있는 경우 frequency와 amplitude 모두에서 beating effect 발생
- 그 결과 입력 신호에 물리적으로 존재하지 않은 주파수 특징 나타남

Significance of Clutter

- 분석된 데이터로부터 신호가 직접 재합성 된다면, 부가적 주파수 요소와 beating effect의 문제 야기하지 않음
- Beating과 다른 예외적인 것들은 signal이 직접 재합성 될 경우 무해하지만 스펙트럼을 시각적으로 자세히 살펴볼 경우 애매한 문제 발생
- clutter는 이러한 분석의 인공적 산물 의미
 Dolson(1983)과 Strawn(1985a): 악기의 tone 분석에 있어 clutter의 중요성 평가
- Gerzon(1991): time과 frequency의 resolution을 향상시킨 super resolving 이론 제시

Alternative Resynthesis Techniques

- 재합성의 일반적 기법을 대체할 두 가지 다른 방법
  ① ABS/OLA(Analysis-by-synthesis/overlap-add)
   : 향상된 resolution을 제공하는 방식으로 더욱 막강한 변형 제공
   error의 분석 절차를 통합하며, 이 진행은 원 신호와 재합성 신호를 비교한 것
   에러가 주어진 threshold 이상인 경우, 분석 프레임에서 amp, freq, phase를 원 소리에 더욱 가깝게 조절. 이러한 진행 과정은 신호가 어느 정도 정확하게 재합성 될 때까지 반복적으로 발생
   이 방식은 일반적인 overlap-add 방식보다 훨씬 정확하게 attack transient, inharmonic spectra, vibrato와 같은 효과 등을 조정
   더욱 강력한 음악적 변환 가능
  ② FFT^{-1}
   : 재합성에 있어 매우 빨라진 속도 제공
   실시간 운영을 위한 방식으로 overlap-add와 oscillator bank resynthesis의 특별한 혼합으로 만들어짐
   inverse FFT에 의해 재합성 되는 이유에서 FFT^{-1}라고 함
   미리 계산된 oscillator bank resynthesis data에서 출발
   이후 이 데이터는 효과적인 알고리듬에 의해 overlap-add 모델로 변환

참고문헌 및 사이트
Curtis Roads, *The Computer Music Tutorial*, pp. 550～563
http://korea.maxim-ic.com/glossary/definitions.mvp/term/frequency_bin/gpk/136
http://en.wikipedia.org/
http://ccrma-www.stanford.edu/~jos/GlobalJOSIndex.html
http://www.clecom.co.uk/science/autosignal/screenshots009.html
http://eceserv0.ece.wisc.edu/~sethares/htmlRT/vocoders/phasevocoder.html

작성자
조원주 (박사과정, 4기)