

# 청각계의 시간 및 주파수 특성을 고려한 VFR-STFT 알고리즘 제안

정 혁, 이 정 권

한국과학기술원 기계공학과 소음 및 진동제어 연구센터

## Use of a New Algorithm of the STFT with Variable Frequency Resolution for the Time-Frequency Auditory Model

Hyuk Jeong and Jeong-Guon Ih  
 NoViC, Department of Mechanical Engineering, KAIST  
 (e-mail:ihih@sorak.kaist.ac.kr)

### 요약

본 연구에서는 청각계의 시간 및 주파수 특성을 고려한 과도음의 시간-주파수 신호해석 기법인 VFR-STFT (STFT with Variable Frequency Resolution)을 제안하고자 한다. VFR-STFT은 downsampling과 FFT를 반복적으로 수행하여 주파수 대역에 따라 주파수 및 시간 분해능이 청각계의 특성과 유사한 기존의 VFR-FFT에 그 뿌리를 두고 있다. 그러나, 본 연구에서는 기존의 VFR-FFT 알고리즘에 overlap 인자를 도입하여 시간-주파수 해석 결과를 구하고, 2/3-rate resampling에 의해 추가로 구성된 시간-주파수 해석 결과의 일부를 기존의 시간-주파수 해석 결과에 이식시킴으로써 기존의 VFR-FFT가 갖는 overlap과 spectral loss 등의 문제점을 최소화하고자 한다.

### 1. 서론

청각 인지 모델과 같이 청각과 관련된 신호 처리를 위해서는 청각계의 시간 및 주파수 해상도에 대한 특성 파악이 진행 되어야 한다. 비록 지난 수십년에 걸쳐서 인간의 청각계에 관한 연구가 진행 되었지만, 청각계의 시간 및 주파수 해상도는 주파수 대역에 따라 변화하기 때문에 청각계에 의한 과도음의 시간-주파수 해석을 동시에 잘 묘사 할 수 있는 방법을 구원하는 것은 매우 어렵다.

청각계의 시간-주파수 해석 모델로서 쉽게 생각할 수 있는 것은 STFT (Short Time Fourier Transform)이다. 그러나, STFT은 주파수 대역에 관계 없이 일정한 주파수 해상도를 제공하기 때문에 청각계의 시간-주파수 해석 모델로는 적합하지 못하다. VFR-FFT [1]는 downsampling과 FFT를 반복적으로 수행하여 주파수 대역에 따른 주파수 및 시간 분해능이 청각계와 유사한 특성을 제공할 수 있는 시간-주파수 해석 방법이다. 그러나 기존의 VFR-FFT 처리 과정에서 overlap을 고려되지 않고 있으며 downsampling과정에서 스펙트럼의 손실이 우려되기 때문에 이에 대한 개선이 요구된다. 본 연구에서는 기존의 알고리즘에 overlap 인자를 도입하여 시간-주파수 해석 결과를 구하고, 2/3-rate resampling에 의해 추가로 구성된 시간-주파수 해석 결과의 일부를 기존의 시간-주파수 해석 결과에 이식함으로써 기존의 VFR-FFT가 갖는 문제점을 최소화하는 VFR-STFT을 제안하고자 한다.

### 2. VFR-STFT

기존의 VFR-FFT에서는 STFT에서 일반적으로 적용되는 겹침 해석 (window overlaped analysis)을 하지 않으므로 윈도우에 의한 데이터 손실이 발생한다. 이를 최소화하기 위해 적절한 겹침 인자 도입이 요구된다. 여기서 겹침 인자는 해석 윈도우 길이에 대한 반복적인 해석 윈도우의 겹침 비율을 의미하며 적절한 겹침 인자란 전 시간 축에서의 가중 함수를 일정하게 하는 겹침 인자를 뜻한다. Hanning 윈도우를 이용하여 해석하는 경우에는 윈도우 길이의 2/3 정도가 겹칠 때 전 시간 축에서의 가중 함수가 일정하게 되며 [2] 계산 효율을 고려할 때 2/3의 겹침 인자가 가장 적절하다고 판단된다.

그림 1은 전 시간 축에서 가중 함수를 일정하게 만드는 과정을 간략하게 나타낸 것으로서 전체 단계가 3일 때의 알고리즘이다.  $q$ 와  $T_j$ 를 각각 윈도우의 겹침 인자 및  $j$ 번째 단계에서의 윈도우 길이로 정의할 때 해석 윈도우의 이동  $\Delta t_j$ 은  $(1-q)T_j$ 로 나타낼 수 있다. 만일 첫번째 시간 블록에서의 해석 윈도우의 중심을 0 초로 설정하면  $j$ 번째 단계에서  $j$ 번째 시간 블록에서의 윈도우 중심  $t'_j$ 는 다음과 같이 나타낼 수 있다:

$$t'_j = (j-1) \cdot \Delta t_j = (j-1) \cdot (1-q) \cdot T_j, j = 1, 2, 3, \dots \quad (1)$$

신호의 구간을 이동하면서 구한  $N_{FFT}$ 개의 스펙트럼 중에서  $N_{FFT}/4+1$ 로부터  $N_{FFT}/2$ 까지의 스펙트럼 샘플을 취하여 그림1의 오른쪽과 같이 각 블록별로 배열한다. 그 다음에는 전체 음향 신호를 downsampling 한다. 그림1에서  $LF$ 는 downsampling시 aliasing을 최소화하기 위한 저역 통과 필터로서 본 연구에서는 8차 elliptic 저역 통과 필터를 사용하였다. 또한 필터링에 의한 위상 왜곡을 방지하기 위하여 zero-phase 디지털 필터링 기법 [3]을 적용하였다.

한편, VFR-STFT의 각 단계가 증가함에 따라서 윈도우의 이동폭도 2배씩 증가하기 때문에 그림 1의 오른쪽과 같이 TFM (Time-Frequency Map)을 표시하기 위한 각 단계에서의 시간 간격이 일정하지 못하게 된다. 일반적으로 TFM과 같은 contour plot을 나타내기 위해서는 각 단계 별로 동일한 시간 간격이 필요하며 각 시각에서의 전체 레벨을 관찰하기 위해서도 각 단계별로 동일한 시간 간격이 되도록 정하는 것이 유리하다. 본 연구에서는 각 단계별로 동일한 시간 간격의 스펙트럼이 나타나도록 주변에 이미 주어졌던 스펙트럼으로부터 선형 보간을 이용하여 스펙트럼이

주어지지 않은 시간 축에서의 값을 구하였다. 이때 선형 보간은 각 파워 스펙트럼을 기준으로 이루어졌다.

VFR-FFT 알고리즘에서는 downsampling 과정에서 anti-aliasing 저역 통과 필터링으로 인하여 필터의 transition 영역에서는 여전히 스펙트럼이 손상되는 문제가 발생한다. 이를 개선하기 위하여 본 연구에서는 그림 2와 같은 알고리즘을 제안한다. 그림 2에서 왼쪽 TFM은 그림 1에서 기술된 알고리즘에 의한 시간-주파수 선도를 나타낸다. 이때 왼쪽 TFM에서 빗금친 영역은 anti-aliasing 저역 통과 필터링에 의해 스펙트럼이 손상된 부분을 의미한다. 반면 최초의 시간-주파수 선도에서 스펙트럼이 손상된 주파수 영역에서 스펙트럼이 손상되지 않은 추가 시간-주파수 선도를 구할 수 있다면 추가된 시간-주파수 선도의 일부를 최초의 시간-주파수 선도에 이식함으로써 새로운 시간-주파수 선도를 구할 수 있으며 이때 스펙트럼의 손상은 최소가 되도록 할 수 있다.

최초의 시간-주파수 선도에 이식하게 될 새로운 시간-주파수 선도를 추가로 구하기 위해서 최초의 샘플링된 신호  $p(t)$ 를 2/3 비율로 리샘플링한 신호  $p'(t)$ 를 이용한다. 2/3 비율로 리샘플링하는 과정은 참고 문헌 [3]에 자세히 기술되어 있다. 추가의 시간-주파수 선도는 최초의 시간-주파수 선도를 얻을 때와 동일하게 그림 1에서 기술된 알고리즘을 이용한다. 이와 같은 과정을 통해 그림 2의 가운데 나타낸 시간-주파수 선도를 얻을 수 있다. 이때 최초의 시간-주파수 선도와 마찬가지로 추가된 시간-주파수 선도에서도 anti-aliasing 저역 통과 필터링에 의해 스펙트럼이 손상된 부분이 빗금친 영역에서와 같이 발생된다. 그렇지만 최초의 시간-주파수 선도와 추가된 시간-주파수 선도에서 스펙트럼이 손상된 주파수 대역이 서로 겹치지 않는다면 스펙트럼이 손상되지 않은 부분만을 추출하여 스펙트럼이 거의 손상되지 않은 새로 구성된 시간-주파수 선도를 구성할 수 있다.

그림 2에서  $f_c^{(0)}$ 부터  $f_N^{(0)}$ 까지, 그리고  $f_c^{(1)}$ 부터  $f_N^{(1)}$ 까지의 주파수 대역은 각각 최초 및 추가 시간-주파수 선도에서 스펙트럼이 손상된 부분을 의미한다. 여기서  $f_c^{(0)}$  및  $f_N^{(0)}$ 는 각각 최초 및 추가 시간-주파수 선도의  $i$ 번째 단계에서 필터의 상계 주파수를 의미하며,  $f_c^{(1)}$  및  $f_N^{(1)}$ 는 각각 최초 및 추가 시간-주파수 선도의  $i$ 번째 단계에서 Nyquist 주파수를 의미한다. 이때  $i$ 번째 단계에서의 Nyquist 주파수  $f_N^{(0)}$ ,  $f_N^{(1)}$ 는 각각 다음과 같이 나타내어진다.

$$f_N^{(0)} = \frac{F_s}{2^{i+1}}, f_N^{(1)} = \frac{F_s}{3 \cdot 2^i} \quad (2.3)$$

이때  $f_c^{(0)}$ 이  $f_N^{(0)}$ 보다 크거나 같고  $f_c^{(1)}$ 이  $f_N^{(1)}$ 보다 크거나 같다고 가정하면 최초 시간-주파수 선도에서 손상된 스펙트럼 대신 손상되지 않은 추가 시간-주파수 선도의 일부를 최초 시간-주파수 선도로의 이식할 수 있으며, 스펙트럼의 손실이 없는 새로운 시간-주파수 선도를 구성할 수 있다. Elliptic 필터의 경우에 차수가 6 이상여 되면 이러한 조건이 만족되는 것을 확인할 수 있었다.

이와 같이 그림 1 및 그림 2의 알고리즘을 통해 윈도우에 의한 데이터 손실과 downsampling 시 anti-aliasing 필터에 의한 스펙트럼 손실이 최소화된 시간-주파수 해석 결과를 얻을 수 있으며, 본 연구에서는 그림 1과 그림 2에 의한 시간-주파수 해석 기법을 VFR-STFT으로 명명하였다.

한편 다음과 같은 예제를 통해 STFT, VFR-FFT 및 VFR-STFT에 의한 시간-주파수 해석 결과를 비교하였다. 예제 해석에 이용된 각 시간-주파수 해석 기법의 주파수 해상도는 그림 3과 같다. 그림에서 임계 대역폭 (critical bandwidth) [4]은 라우드니스 해석을 위한 정각 필터의 대역폭을 JNVE (just-noticeable variation in frequency) [4]은 변조 주파수가 4 Hz인 주파수 변조음으로부터 측정된 정각계의 순음 분해능을 나타낸다. Terhardt 등은 Hanning 윈도우를 사용할 때 1 kHz 미만에서는 스펙트럼 간격이 약 10 Hz 이면 순음의 여부를 판정할 수 있음을 확인하였다 [5]. STFT은 주파수 대역과는 관계 없이 동일한 주파수 해상도를 제공하고 VFR-FFT 및 VFR-STFT은 주파수 대역에 따라 주파수 해상도가 변하기 때문에 약 1 kHz 미만의 저주파 대역에서만 모두 약 10 Hz의 동일한 주파수 해상도를 제공하도록 설정하였다. 또한 STFT이나 VFR-STFT의 경우에는 검출 해석을 위한 검출 인자가 필요한데 본 예제에서는 모두 2/3의 검출 인자를 도입하였다.

예제로서 이용된 신호는 시각에 따라 주파수가 증가하는 정현파 소인 신호  $p_1(t)$ 와 고주파 대역의 반복적인 펄스 신호  $p_2(t)$ 로 구성되어 있다. 최초 2초 동안은 정현파 소인 신호가 이후에는 반복적 펄스 신호가 존재한다. 정현파 소인 신호는 downsampling에 의한 스펙트럼 손실을 파악하기 위하여 사용되었고, 이에 정현파 소인 신호의 음압 레벨은 60dB이며 순간 주파수는 다음과 같다.

$$f(t) = 200 \cdot \exp(2t) \quad (4)$$

고주파 대역의 펄스 신호  $p_2(t)$ 는 중심 주파수  $f_c$ 가 4 kHz, 노출 시간이 30ms 이고 그 주기는 60ms 이다. 이때 펄스 신호의 최대 음압 레벨은 60dB 이다. 본 예제 펄스 신호의 노출 시간 및 발생 주기는 과도음의 라우드니스 모델에서 제안된 시간 간격 2 ms에 비해 매우 넓기 때문에 펄스 각각의 크기 변화가 시각에 의해 인지된다. 따라서 예제의 펄스 신호를 통해서 각 시간-주파수 해석 기법의 시간 해상도가 정각 해석에 적합한지를 간접적으로 확인할 수 있다.

그림 4는 STFT로부터 TFM을 구하고 이로부터 전체 음압 레벨을 나타낸 것이다. 상단은 STFT에 의한 TFM을 음압 레벨로서 나타낸 것이며 하단은 TFM으로부터 구한 전체 음압 레벨이다. 또한 그림 5와 그림 6은 각각 VFR-FFT와 VFR-STFT에 의한 TFM과 전체 음압 레벨을 나타낸 것이다.

그림 4의 STFT로부터 구한 전체 음압 레벨은 주파수 대역에 관계없이 정현파 소인 신호의 전체 음압 레벨과 동일한 결과를 보여주고 있다. 반면 그림 5의 VFR-FFT에 의한 결과는 VFR-FFT 각 스텝 단계에서 스펙트럼의 손실이 발생하는 것을 확인할 수 있다. 이는 downsampling 과정에서의 low-pass filtering에 의해 스펙트럼이 손상되었기 때문이다. 그림 6의 VFR-STFT 경우에는 정현파 소인 신호의 전체 음압 레벨이 0.7 dB 이내로 거의 동일한 결과를 보여주고 있다.

한편 반복적인 펄스에 의한 TFM과 전체 음압 레벨을 관찰하면 그림 4의 STFT의 결과에서는 펄스의 존재를 거의 인지할 수 없다. 이는 저주파 대역에서 다른 시간-주파수 해석법과 동일한 주파수 해상도를 가지기 때문에 정각의 모델에 필요한 충분한 시간 해상도를 확보할 수 없기 때문이다. 그림 5의 VFR-FFT 경우에는 펄스의 존재는 확실하게 파악할 수 있지만 동일한 모양의 펄스임에도 시간축의 위치에 따라 서로 다른 형상의 펄스로 나타나며 그 최대 음압 레벨 또한 차이를 확인할 수 있다. 이는 기존의

VFR-FFT 알고리즘에서 겹침 해석이 이루어지지 않았기 때문에 발생하는 결과이다. 그림 6의 VFR-STFT 경우에는 고주파 대역에서의 펄스들을 서로 구분하여 파악할 수 있으며 전체 음압 레벨에서도 각 펄스에 따라 거의 동일한 형상을 확인할 수 있다.

따라서 전체적인 성능을 비교할 때 VFR-STFT이 기존의 STFT이나 VFR-FFT에 비해 청각계의 시간 및 주파수 분해능과 유사한 특성을 가지고 있기 때문에 과도적인 오디오 신호의 라우드니스 해석을 위한 시간-주파수 해석법으로서 유용하다고 판단된다.

### 3. 결론

본 연구에서는 청각계의 시간 및 주파수 특성을 고려한 과도음의 시간-주파수 해석 기법, 즉 VFR-STFT을 제안하였다. VFR-STFT은 downsampling와 FFT를 반복적으로 수행하여 주파수 대역에 따라 주파수 및 시간 분해능을 청각계의 특성과 유사하게 적용할 수 있다. 따라서 VFR-STFT은 기존의 방법에 비해 청각계와 유사한 시간 및 주파수 분해능을 제공하고 있다. 비록 VFR-STFT 기본 개념이 기존의 VFR-FFT에 기초하고 있지만, 기존의 VFR-FFT 알고리즘에서 고려되지 않은 overlap 인자를 도입하고 2/3 rate resampling에 의해 추가로 구성된 음향 신호의 시간-주파수 해석 결과 일부를 최초의 시간-주파수 해석 결과에 이

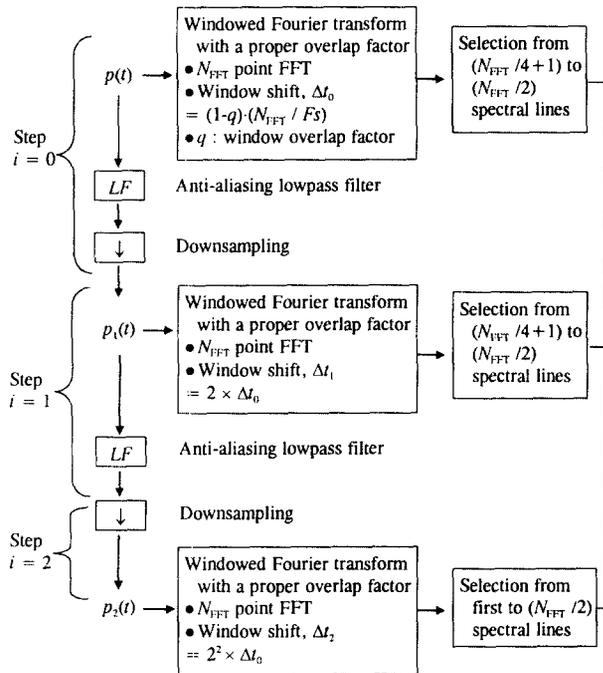


Fig. 1. Block diagram showing the procedure to provide completely uniform overall weighting function in the VFR-STFT by introducing the proper overlap factor to the analysis window.

식시킴으로써 기존 VFR-FFT의 문제점을 최소화 하였다. 또한 본 연구에서는 downsampling과정의 저역 통과 필터링시 zero-phase 디지털 필터링 기법을 이용함으로써 필터링에 의한 위상 왜곡을 최소화하였다.

### 참고문헌

- [1] Anon., *Binaural Analysis System Manual*, Appendix B (HEAD-Acoustics GmbH, Herzogenrath, 1996).
- [2] R.B. Randall, *Frequency Analysis*, 3rd ed., Chap. 4 (K. Larsen & Son A/S, Denmark, 1987).
- [3] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*, Chaps. 6 and 11, (Prentice-Hall, New Jersey, 1989).
- [4] E. Zwicker and H. Fastl, *Psychoacoustics, Facts and Models*, (Springer-Verlag, Berlin, 1990).
- [5] E. Terhardt, G. Stoll, and M. Seewann, "Algorithm for Extraction of Pitch and Pitch Saliency from Complex Tonal Signals," *J. Acoust. Soc. Am.*, vol. 71, pp. 679-688 (1982).

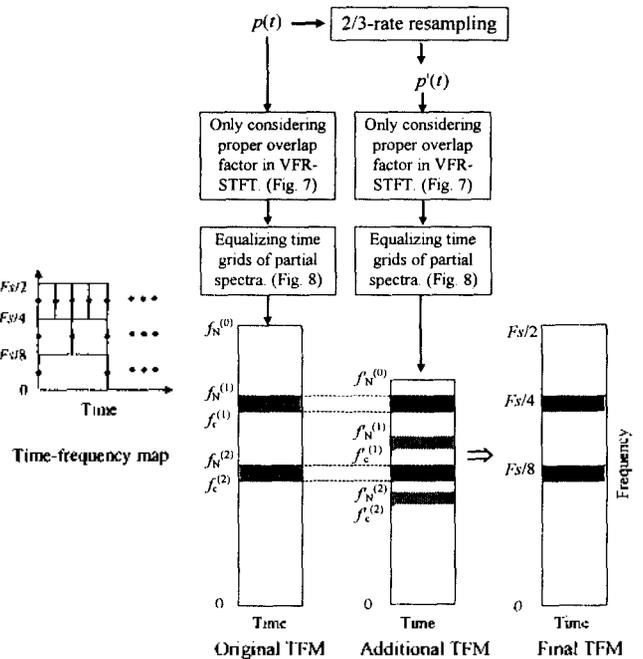


Fig. 2. Schematic procedure to minimize spectrum-impaired frequency region by the spectrum transplant. ■■■■, impaired spectra due to the anti-aliasing filter; ■■■■, transplanted spectra to the original TFM.

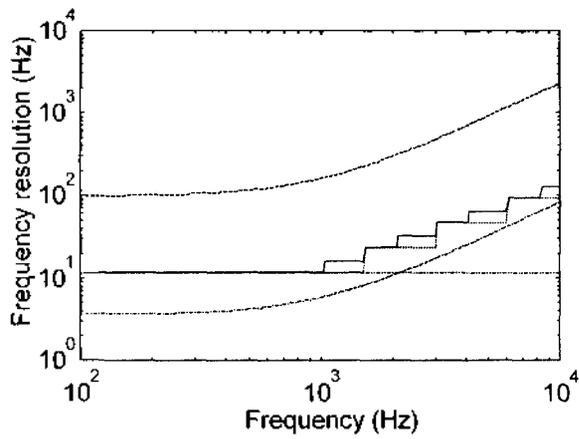


Fig. 3. Comparison of frequency resolutions: ..... , critical bandwidth; - . - . , JNVF; - - - - , STFT; - - - - , VFR-FFT; ———— , VFR-STFT.

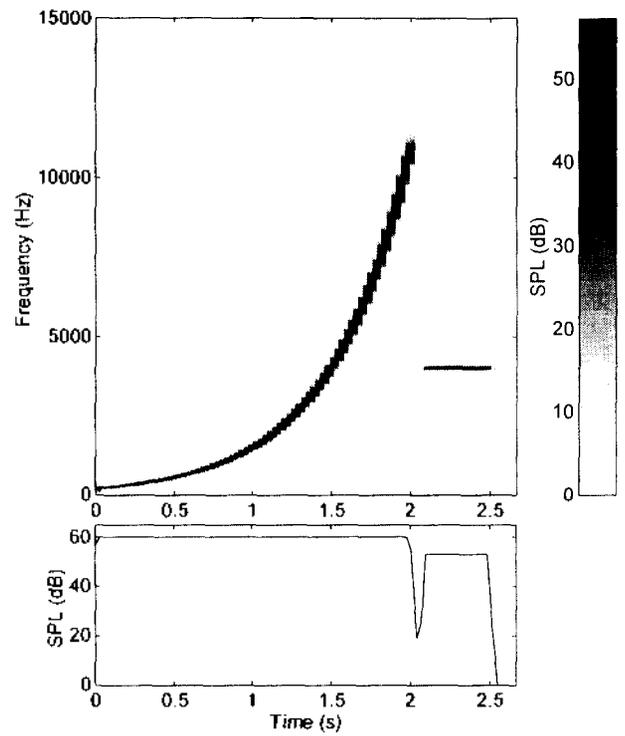


Fig. 4. TFM and the overall sound pressure level of the test signal obtained by the STFT.

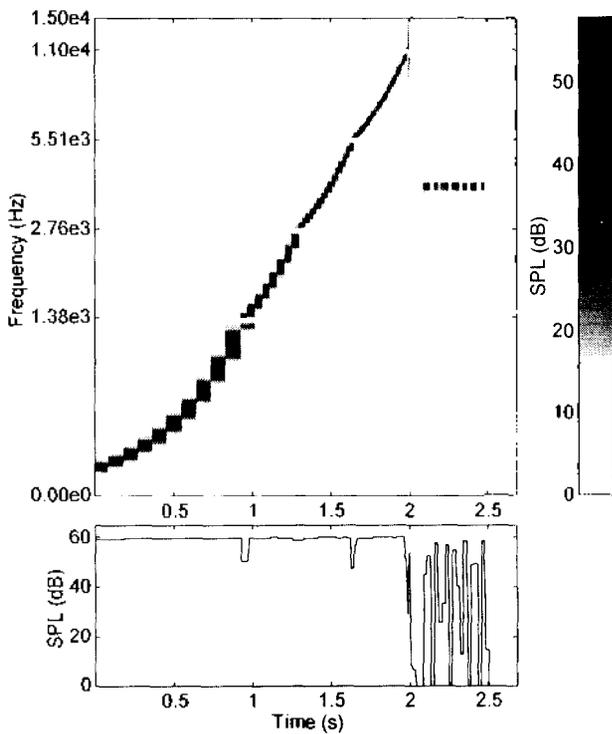


Fig. 5. TFM and the overall sound pressure level of the test signal obtained by the VFR-FFT.

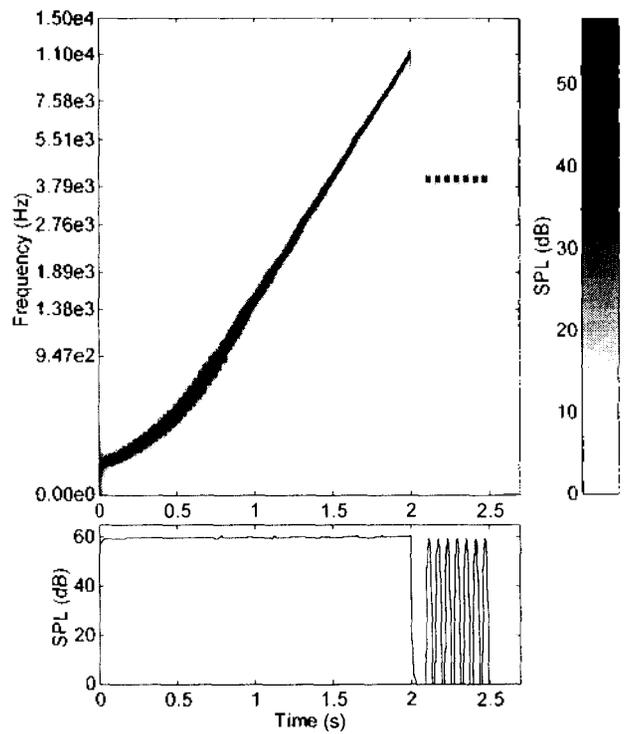


Fig. 6. TFM and the overall sound pressure level of the test signal obtained by the VFR-STFT.