FM 변조된 형태의 Kernel을 사용한 음성신호의 시간-주파수 표현 해상도 향상에 관한 연구

On Improving Resolution of Time-Frequency Representation of Speech Signals Based on Frequency Modulation Type Kernel

이 희 영* · 최 승 호**

Heyoung Lee · Seungho Choi

ABSTRACT

Time-frequency representation reveals some useful information about instantaneous frequency, instantaneous bandwidth and boundary of each AM-FM component of a speech signal. In many cases, the instantaneous frequency of each component is not constant. The variability of instantaneous frequency causes degradation of resolution in time-frequency representation. This paper presents a method of adaptively adjusting the transform kernel for preventing degradation of resolution due to time-varying instantaneous frequency.

The transform kernel is the form of frequency modulated function. The modulation function in the transform kernel is determined by the estimate of instantaneous frequency which is approximated by first order polynomial at each time instance. Also, the window function is modulated by the estimated instantaneous frequency for mitigation of fringing effect. In the proposed method, not only the transform kernel but also the shape and the length of the window function are adaptively adjusted by the instantaneous frequency of a speech signal.

Keywords: Time-frequency Representation, Time-varying Harmonic Structure, Instantaneous Frequency, Instantaneous Amplitude, Resolution, AM-FM Component

1. 서 론

시간-주파수 표현 (time-frequency representation)은 시변 특성을 가지는 신호들의 해석에 적합하며, 음성신호 분해(decomposition), 음질향상(speech enhancement), 음질평가, 음성합성, 음성코딩, 음성인식, 음성의학 등의 음성신호처리 분야에서 유용하게 활용된다[1][13][19][21]. 음성과 같은 다중 성분 신호(multi-component signal)를 시간-주파수 평면에서 해석할 경우, 시간 영역이나

* 서울산업대학교 공과대학 제어계측공학과
** 서울산업대학교 공과대학 전자정보공학과

Fourier 주파수 영역 만에서는 관찰하기 어려운 스펙트럼의 시간-주파수 분포, 각 성분들의 시간적 변화 양상 등 여러 가지 성질 및 특성을 파악할 수 있다[2][3][4][5][6][15][16]. 즉, 각 성분들의 순간 주파수(instantaneous frequency)의 궤적, 순간 진폭(instantaneous amplitude)의 강도, 순간 대역폭(instantaneous bandwidth)의 경계 등과 같은 정보를 얻을 수 있다[9][10][11][16][20].

신호의 시간-주파수 표현의 해상도는 근본적으로 시간, 주파수에 대한 불확정성 원리(uncertainty principle)에 지배를 받으므로 시간 축과 주파수 축에서 동시에 높은 해상도를 얻을 수 없다[17]. 그러나 다중 성분 신호가 일정한 진폭을 갖는 정현파의 선형 결합으로 이루어져 있을 경우 Short-Time Fourier 변환에 기반을 둔 방법들은 신호들에 대해 매우 높은 해상도의 시간-주파수 표현을 제공한다[7][16][19]. 이것은 변환 커널(transform kernel)의 구조가 정현파의 순간 주파수를 잘 반영하는 형태로 되어 있기 때문이다[16].

시간-주파수 표현의 해상도를 떨어뜨리는 이유로는 각 성분들의 시변 특성을 들 수 있다[18][22]. 많은 경우 다중 성분 신호는 각 성분들의 순간 진폭, 순간 주파수들이 일정하지 않고 시간에 따라 변하는데 Short-Time Fourier 변환에 기반을 둔 방법들은 시변 특성을 제대로 반영하지 못하는 커널 구조 때문에 시간-주파수 표현의 해상도를 높이는데 한계가 있다[18][22]. 즉 각 성분들의 시변 특성으로 인해 주파수 축 및 시간 축으로 간섭이 일으켜서 시간-주파수 표현 해상도가 떨어진다[18][22]. 한편 Wigner-Ville 분포(distribution)와 같은 비선형 변환(bilinear transform)은 단일 성분 신호에 대한 시간-주파수 표현 해상도가 Short-Time 변환에 비하여 높으나 다중 성분 신호의 경우 성분들 사이의 간섭 효과가 심하여 음성신호와 같은 다중 성분 신호의 시간-주파수 표현에는 사용하기가 어렵다[5].

음성신호의 시간-주파수 표현에는 스펙트로그램(spectrogram)이 주로 사용되는데 해상도를 높이기 위하여 창 함수(window function)의 길이와 모양을 신호에 적응적으로 변화시키는 방법들이 제안 되었다[18][22]. ACKD(adaptive cone-kernel distribution) 방법의 경우는 커널 함수를 주어진 신호의 특성에 적응적으로 변화시킨다[22]. ACKD 방법에서 커널 함수는 주어진 신호에 적응적으로 변하는 함수와 Fourier 변환의 커널 함수의 결합으로 이루어져 있다. ASTFT(adaptive short-time Fourier transform) 방법[18]은 커널 함수의 모양은 변화시키지 않고 창 함수의 길이를 신호의 특성에 적응적으로 변화시킨다. 이 방법들을 사용하여 시간-주파수 표현을 구할 경우, 신호의 순간 주파수가 시간적으로 변하지 않는 일정한 부분에서는 해상도가 높아지는 특성을 가지고 있다. 그러나 ACKD와 ASTFT 모두 변환 커널이 $\exp(j\omega t)$ 함수를 기반으로 하고 있으므로 신호의 순간 주파수의 변화가 심한 부분에서는 해상도를 증진시키는데 한계가 있다. 예를 들어 선형 첩 신호(*linear chirp signal*)와 같은 시변 순간 주파수를 갖는 신호는 Fourier 스펙트럼이 집중되지 않고 분산되므로 ADKC와 ASTFT 방법에 의한 시간-주파수 표현은 순간 주파수가 변하는 부분에서 해상도가 떨어진다.

본 논문에서는 음성 신호의 시간-주파수 표현 해상도를 높이기 위하여 변환 커널을 순간 주파수에 적응적으로 변화시키는 시간-주파수 표현 방법을 제안한다. 제안된 방법에서 커널 함수는 음성 신호의 제 1 순간 주파수에 따라 적응적으로 변화하는데, 이와 같이 할 경우 순간 주파수가 변함으로 인해 발생하는 해상도 저하 효과를 방지할 수 있다. 본 논문은 다음과 같이 구성된다. 2 장에서 음성 신호의 모음 부분에 대한 AM-FM 모델을 다루었고, 3 장에서 시간-주파수 표현을 위한

커널 함수의 설계 방법 다루었다. 4 장에서는 제안된 방법을 음성신호에 적용하고 실험 결과를 분석하며, 5 장에서 결론을 맺는다.

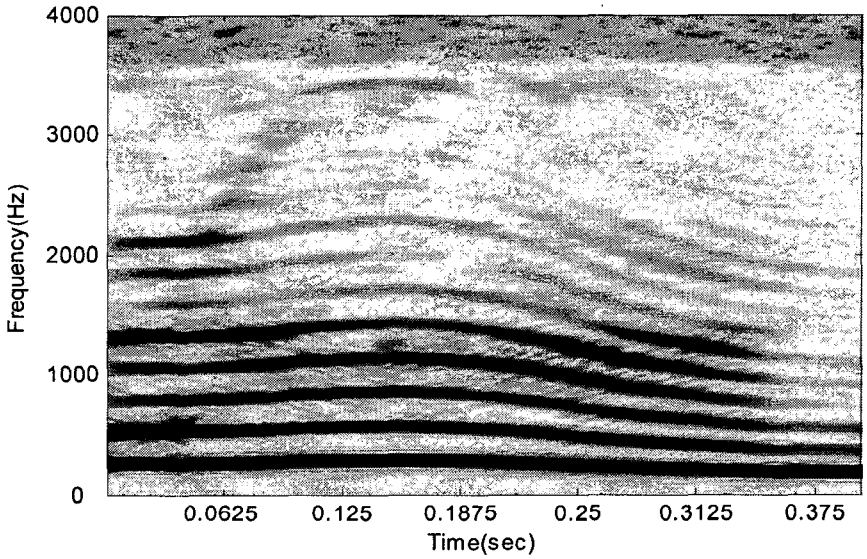


그림 1. 음성 신호의 스펙트로그램

2. 음성 신호의 AM-FM 모델

음성신호의 모음부분은 다중 성분 신호이며 다음과 같이 AM-FM 신호의 합으로 모델링할 수 있다[2][9][12][14][15].

$$x(t) = \sum_{n=0}^{N-1} a_n(t)\cos(\phi_n(t)) + N(t) \tag{1}$$

여기서 $a_n(t)$는 n 번째 AM-FM 성분의 순간 진폭(instantaneous amplitude)이고 $\phi_n(t)$는 n 번째 AM-FM 성분의 순간 위상(instantaneous phase)이다. 식 (1)은 음성 신호를 준 주기적 신호(quasi-periodic signal)인 AM-FM 성분들과 잡음을 포함한 비 주기적인 신호 $N(t)$ 나누었다 [[14][15]. 일반적으로 음성 신호 모음의 에너지는 주로 AM-FM 성분에 집중되므로 $N(t)$의 에너지는 상대적으로 작다. 각 성분의 순간 주파수(instantaneous angular frequency) $\omega_n(t)$는 순간 위상의 시간 미분으로 정의되며 다음과 같다[1][5][16].

$$\omega_n(t) = D\phi_n(t) \quad \text{rad/s} \tag{2}$$

여기서 $D=d/dt$이다.

<그림 1>은 20 대 여성이 발성한 영어 문장의 모음 부분을 8 kHz로 표본화한 신호의 스펙트로그램이다. 창 함수의 길이는 30 msec로 하였다. 음성신호 파형은 인간의 음성 발생 시스템의 동적 특성 및 경계 조건에 의해 결정되는데, <그림 1>에서 서로 다른 순간 주파수들 간의 간격이 일정하지 않음을 알 수 있다. 즉 <그림 1>에서 고조파 순간 주파수들은 기본 순간 주파수 $\omega_0(t)$에 비례한다. <그림 1>에서 순간 주파수의 변화가 심한 부분, 즉 시간 구간 $0.1<t<0.12$ 에서 고조파 순간 주파수들은 기본 순간 주파수에 비례하는 것을 관찰 할 수 있다. 즉

$$\omega_n(t)\simeq b_n\omega_0(t) \tag{3}$$

여기서 b_n , $n=0,1,2,\cdots,N-1$는 상수이다. 비교적 짧은 시간 구간에서는 식 (3)의 $\omega_0(t)$를 1차 다항식으로 근사화 시킬 수 있다. 즉 t_0에서

$$\begin{aligned}\omega_n(t)&\simeq b_n(\alpha(t-t_0)+\beta),\\&=\beta b_n(\frac{\alpha}{\beta}(t-t_0)+1),\quad t_0-\delta\le t\le t_0+\delta\end{aligned} \tag{4}$$

여기서 α와 β는 상수이고 δ는 근사화 범위를 나타내는 상수이다.

일반적으로 각각의 성분은 순간 주파수와 순간 진폭 모두가 시간에 따라 변하는 특성을 갖는다. 그러나 비교적 짧은 구간에서는 각 성분의 순간 진폭은 상수로 근사화 할 수 있다. 그러므로 식 (1), (2), (3), (4)로부터 짧은 구간에서 음성신호의 모음 부분은 다음과 같이 근사적으로 표현 된다. 순간 주파수는 순간 위상의 시간 미분이므로 식 (2)와 (4)로부터 음성 신호를 t_0에서 근사화 시키면 다음과 같은 수식을 얻을 수 있다.

$$x(t)\simeq\sum_{n=0}^{N-1}a_n\cos(b_n(\frac{1}{2}\alpha t^2-\alpha t_0t+\beta t)+c_n)+N(t),\quad t_0-\delta\le t\le t_0+\delta \tag{5}$$

여기서 a_n는 상수이고 순간 진폭을 근사화한 것이다. c_n은 적분 상수로 위상 천이와 관련된 상수이다. 식 (5)는 짧은 구간에서는 음성 신호를 일정한 진폭을 갖는 선형 첩 신호(linear chirp signal)의 합으로 근사화 시킬 수 있음을 의미한다.

3. 시간-주파수 표현을 위한 변환 커널의 설계

선형 첩 신호와 같이 변하는 순간 주파수를 가지고 있는 신호는 일정한 순간 주파수를 갖는 정현파와는 달리 Fourier 주파수 영역에서 spectral support가 넓게 퍼진다. 그러므로 음성 신호를 Short-Time Fourier 변환을 사용하여 시간-주파수 표현을 통해 관찰할 경우 순간 주파수가 변하는 부분에서는 해상도가 저하된다. 각 성분들의 순간 주파수가 변하는 음성 신호의 시간-주파수 표현 해상도 저하를 방지하기 위해서는 순간 주파수에 적응적으로 대응하면서 변화할 수 있는 구조의 커널이 필요하다.

3.1 변환 커널 설계

식 (1) 또는 (5)와 같이 시변 순간 주파수를 갖는 신호의 시간-주파수 표현을 위해 확장 Fourier 변환(extended Fourier transform) 및 역 변환을 다음과 같이 정의하자[4][8].

$$F[x(t), g(t)] = \int_{-\infty}^{\infty} x(t) \frac{1}{g(t)} \exp(-j\omega_g \int_0^t \frac{1}{g(\zeta)} d\zeta) dt$$
$$= X(\omega_g) \tag{6}$$

$$F^{-1}[X(\omega_g), g(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega_g) \exp(j\omega_g \int_0^t \frac{1}{g(\zeta)} d\zeta) d\omega_g$$
$$= x(t) \tag{7}$$

여기서 $g(t)$는 양의 함수 이고 ω_g는 주어진 함수 $g(t)$에 대응하는 확장 Fourier 주파수 변수이다. 여기에서 ω_1은 Fourier 주파수 변수이다. 확장 Fourier 변환은 시변 순간 주파수를 갖는 신호를 확장 Fourier 주파수 영역에서 국소화(localization) 시킬 수 있다. 다음과 같은 Dirac 충격 함수를 고려하자.

$$X(\omega_g) = \delta(\omega_g - \omega_0)$$

식 (7)에 의하면 $x(t) = F^{-1}[X(\omega_g), g(t)]$는 다음과 같이 된다.

$$x(t) = \frac{1}{2\pi} \exp(j\omega_0 \int_0^t \frac{1}{g(\tau)} d\tau)$$

이것은 시변 순간 주파수를 갖는 시간 영역 신호의 순간 주파수를 알고 있을 경우 확장 Fourier 주파수 영역에서 완전 국소화 시킬 수 있음을 의미한다. 확장 Fourier 변환에 대한 이와 같은 성질을 이용하면 시변 순간 주파수를 갖는 신호의 시간-주파수 표현 해상도를 향상시킬 수 있는 변환을 설계할 수 있다.

일반적으로 주어진 신호의 시변 순간 주파수를 전 시간 영역에서 정확히 추정하는 것은 매우 어렵다. 그러나 주어진 시점에서 순간 주파수를 선형 근사화 시켜 추정 값을 얻는 것은 비교적 쉽게 할 수 있으므로 다음과 같이 변환을 정의하자.

$$X_a(t,\omega_g)=\int_{-\infty}^{\infty}x(\tau)w(\tau-t)\ker_a(\tau-t)d\tau \tag{8}$$

여기서 $x(\tau)$는 신호이고, $w(\tau)$는 구간 $-\tau_-<\tau<\tau_+$ 에서만 정의되고 나머지 구간에서는 영인 창함수(window function)이다. τ_-와 τ_+는 양의 상수이며 이것들을 결정하는 방법은 이 절의 끝 부분에서 다룬다. $\ker_a(\tau-t)$는 다음과 같은 변환 커널이다.

$$\ker_a(\tau-t)=\frac{1}{g(\tau-t)}\exp(-j\omega_g(\int_0^{\tau}\frac{1}{g(\zeta-t)}d\zeta)) \tag{9}$$

여기서 $1/g(\tau)$는 변환 커널의 주파수 변조 함수이며 다음과 같이 선택하였다.

$$1/g(\tau)=a\tau+1,\quad -W/2<\tau<W/2$$

a는 결정해야 할 파라미터이고 충분히 작아서 구간 $-W/2<\tau<W/2$ 에서 $1/g(\tau)$는 양이다. 그러므로 시간 t에서 변환 커널은 다음과 같다.

$$\ker_a(\tau-t)=(a(\tau-t)+1)\exp(-j\omega_g(\frac{1}{2}a\tau^2-at\tau+\tau)) \tag{10}$$

예를 들어 $t=0$인 경우를 고려해 보자. 이때, 변환 커널은 다음과 같다.

$$\ker_a(\tau)=(a\tau+1)\exp(-j\omega_g(\frac{1}{2}a\tau^2+\tau))$$

주어진 신호 $x(t)$에 대한 시간-주파수 표현 $\overline{X}(t,\omega_g)$은 다음과 같이 정의된다.

$$\overline{X}(t,\omega_g)=|X_{a^*}(t,\omega_g)| \tag{11}$$

여기서 a^*는 다음과 같은 $P(a)$를 극대로 하는 a의 값이다.

$$P(a)=\max_{\omega_g}|X_a(t,\omega_g)|$$

식 (11)은 다음과 같은 의미를 갖는다. 주어진 시간 시점에서 창 함수가 곱해진 신호의 스펙트럼에서 ω_g에 대한 최대치를 구하고 이 최대치를 극대로 하는 a를 구한다. 이때의 스펙트럼을 주어진 시간 시점에서의 단 구간(short term) 스펙트럼으로 선택하여 주어진 신호의 시간-주파수 표현을 얻는다. 음성 신호의 경우 윈도우 창 함수의 구간은 수십 msec 정도이고 a는 너무 크지 않으므로 창 구간에서 함수 $1/g(\tau-t)$가 고려하고 있는 구간에서 양이다. 다음은 창 함수를 결정하는 방법을 고려하자.

스펙트로그램에서는 Hanning 창 함수가 주로 사용되는데 식 (8)에서 Hanning 창 함수를 직접 사용할 경우 창 함수에 의한 스펙트럼 퍼짐 현상이 나타난다. 이와 같은 퍼짐 현상을 방지하기 위하여 다음과 같은 변조된 Hanning 창 함수를 사용한다.

$$\omega(\tau)=0.5(1+\cos(2\pi m(\tau)/W)),\quad -\tau_-<\tau<\tau_+ \tag{12}$$

여기서 창 함수는 $\tau_-<\tau<\tau_+$인 영역에서만 정의 되고 그 외의 영역에서는 영이다. W는 변조되지 않은 Hanning 창 함수의 길이를 나타내는 양의 상수이다. 또한 $m(\tau)$는 다음과 같다.

$$\begin{aligned} m(\tau)&=\int_0^{\tau}\frac{1}{g(\zeta)}\,d\zeta \\ &=\int_0^{\tau}(a\zeta+1)d\zeta \end{aligned} \tag{13}$$

그리고 $\tau_-=m^{-1}(-W/2)$이며 $\tau_+=m^{-1}(W/2)$이다. 변조된 창 함수의 길이는 W의 함수이다.

3.2 제안된 변환 Kernel의 해상도 향상 과정 분석

제안된 변환 Kernel이 어떻게 음성신호의 해상도를 향상 시키는가를 분석해보자. 식 (5), (8), (10), (11) 그리고 (12)를 사용한다. 그리고 편의상 $t=t_0=0$이고 $N(t)=0$이라고 가정하자. 또한 식 (5)의 신호 $x(t)$에 대하여 식(11)을 사용하여 a가 α/β에 동조되었다고 가정하자. 즉 $a^*=\alpha/\beta$라고 가정하자. 또한 시간-주파수 표현의 크기만 구하는 것이므로 $c_n=0$으로 놓아도 무방하다. 식 (8), (10), (11) 그리고 (12)로부터 식 (5)의 $t=0$에서의 시간-주파수 표현 $\overline{X_a}(0,\omega_g)$는 다음과 같이 표현된다. 즉

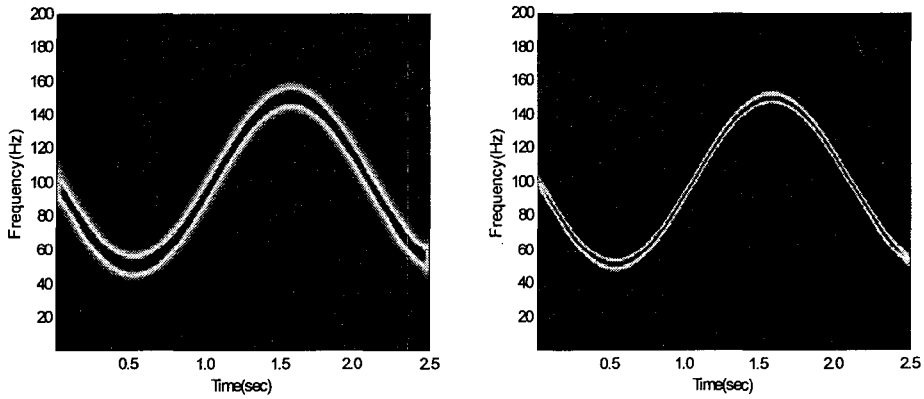
$$\begin{aligned} \overline{X}(0,\omega_g)&=|X_{\alpha/\beta}(0,\omega_g)| \\ &=\left|\int_{-\tau_-}^{\tau_+}x(\tau)w(\tau)\ker_{\alpha/\beta}(\tau)d\tau\right| \\ &=\left|\int_{-\tau_-}^{\tau_+}\sum_{n=0}^{N-1}a_n^0\cos(\beta b_n(\frac{1}{2}\frac{\alpha}{\beta}\tau^2+\tau))0.5(1+\cos(2\pi m(\tau)/W)\right| \end{aligned}$$

$$(\frac{\alpha}{\beta}\tau+1)\exp(-j\omega_g(\frac{1}{2}\frac{\alpha}{\beta}\tau^2+\tau))d\tau| \quad (14)$$

여기서 $m(\tau)=\int_0^{\tau}(\frac{\alpha}{\beta}\zeta+1)d\zeta$ 이다. 변수 치환을 위하여 $\eta=m(\tau)$로 놓자. 그러면 식 (14)는 다음과 같이 표현된다.

$$\overline{X}(0,\omega_g))=|\int_{-W/2}^{W/2}\sum_{n=0}^{N-1}a_n^0\cos(\beta b_n\eta)0.5(1+\cos(2\pi\eta/W))\exp(-j\omega_g\eta)d\eta|$$
$$=|\sum_{n=0}^{N-1}\int_{-W/2}^{W/2}a_n^0\cos(\beta b_n\eta)h(\eta)\exp(-j\omega_g\eta)d\eta| \quad (15)$$

여기서 $h(\eta)=0.5(1+\cos(2\pi\eta/W))$는 Hanning 창 함수이다. 식 (15)는 변환 (11)을 사용할 경우 선형 첩 신호들의 시간-주파수 표현은 단순히 대응하는 정현/여현 신호들의 스펙트로그램으로 나타내짐을 의미한다. 이것은 선형 첩 신호의 시간-주파수 표현에서 Kernel의 주파수 변조 함수를 신호의 순간 주파수에 동조시킬 경우 스펙트럼이 순간 주파수의 시간 변화에 의해 퍼지는 효과를 줄일 수 있음을 의미한다. 또한 여현 신호이므로 W를 늘일 경우 시간-주파수 표현의 주파수 해상도를 높일 수 있다. 임의로 변하는 순간 주파수를 갖는 신호의 경우 W를 순간 주파수가 1 차 다항식으로 근사화가 허용되는 범위까지 늘일 수 있다. 즉 W를 변화시켜 식 (12)의 변조된 Hanning 창 함수의 길이를 변화시킬 수 있다.



(a) (b)

그림 2. 시간-주파수 표현의 해상도 비교. (a) 스펙트로그램, (b) 제안된 방법

3.3 해상도 향상 성능 비교

<그림 2>는 시변 순간 주파수를 갖는 신호 $x(t)=\cos(200\pi t+100\pi\cos(3t)/3)$의 시간-주파

수 표현의 해상도 비교를 보여준다. <그림 2>의 (a)는 길이가 150 ms인 Hanning 창 함수를 사용하여 얻은 스펙트로그램이다. <그림 2>의 (b)는 제안된 방법으로 얻은 시간-주파수 표현이다. <그림 2>의 비교로부터 제안된 방법의 해상도가 상당히 향상되었음을 알 수 있다. <그림 2>의 (b)는 식 (11)을 사용해 얻었으며 a^*를 구하기 위하여 최급 상승 방법(steepest ascent method)이 사용되었다. <그림 2>의 (b)에서 $W=0.4$ msec가 사용되었다. 이때 사용된 창 함수 $\omega(\tau)$의 a에 따른 변화를 <그림 3>에 나타냈다. <그림 3>에서 순간 주파수가 일정하지 않을 경우 창 함수가 좌우 비대칭임을 알 수 있다. <그림 4>는 식 (11)에 의하여 구한 a^*를 보여준다.

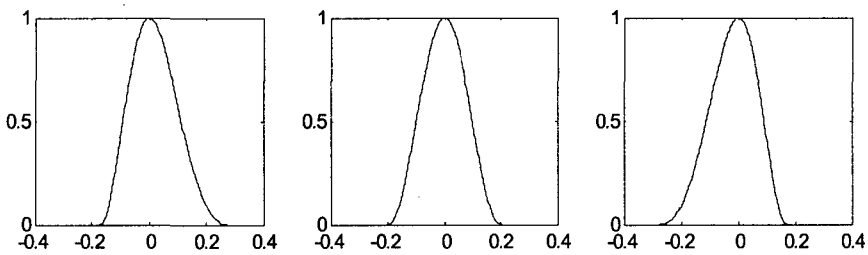


그림 3. 창 함수의 예. (a) $a=-2$, (b) $a=0$, (c) $a=2$

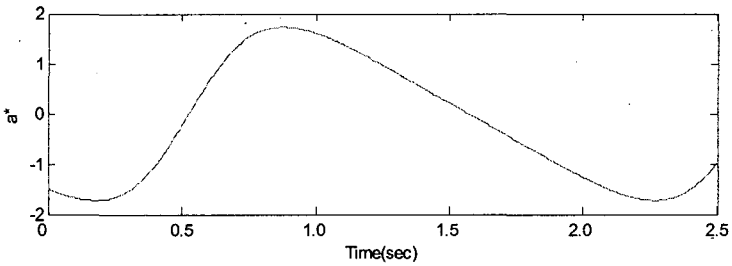
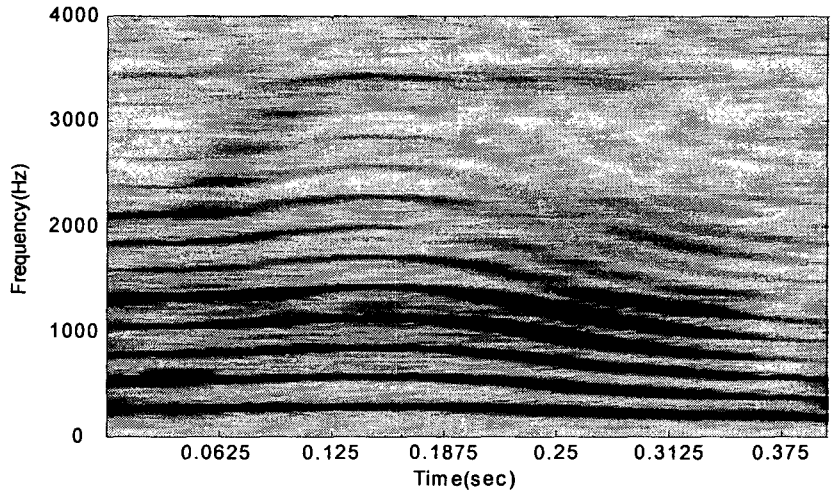


그림 4. 식 (11)의 a* 값의 변화

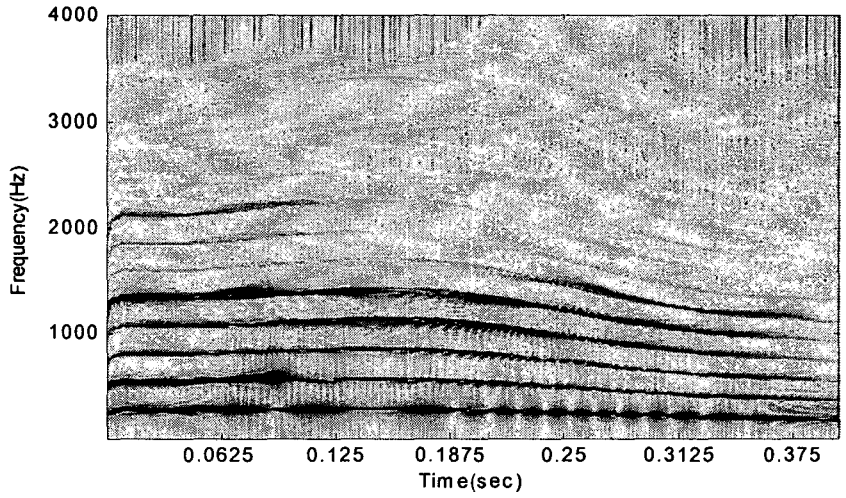
4. 실 험

<그림 5>는 일반적인 스펙트로그램과 본 연구에서 제안한 방법에 의한 시간-주파수 표현들을 비교한 것이다. <그림 5>의 (a)에서 창 함수의 길이를 64 msec로 하였다. <그림 5>의 (a)의 경우 창 함수의 길이를 30 msec로 한 <그림 1>과 비교하면 순간 주파수가 변하는 부분에서 간섭 현상이 있음을 관찰할 수 있다. <그림 5>의 (b)와 (c)는 $W=64$ msec로 한 경우인데, 식 (10)의 커널 함수에서 기울기 a를 신호의 순간 주파수에 적응적으로 변화시켜 얻은 시간-주파수 표현이다. <그림 5>의 (b)와 (c)에서와 같이 제안된 방법은 창 함수 길이가 늘어나서 발생하는 시간 축으로의

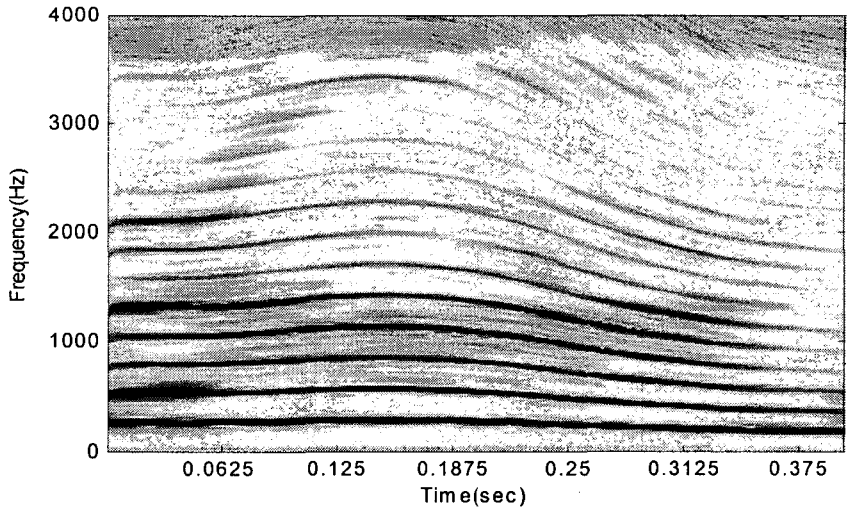
간섭현상을 제거할 수 있다. 일반적으로 창 함수의 길이가 늘이나면 주파수 축 해상도는 증가하지만 시간 축 해상도는 감소한다. 그러나 제안된 방법은 시변 순간 주파수를 갖는 신호에 대하여 창 함수의 길이를 늘려도 시간 축 해상도는 저하되지 않고 주파수 축 해상도를 향상시키는 효과를 거둘 수 있다. <그림 5>의 (b)는 변조되지 않은 창 함수를 사용하여 얻은 시간-주파수 표현으로서 주파수축의 간섭현상을 관찰 할 수 있다. <그림 1> 및 <그림 5>의 (a), (b)와 비교하여 변조된 창 함수를 사용한 <그림 5>의 (c)의 결과를 보면, 기본 순간 주파수와 고조파 순간 주파수들에 대한 시간-주파수 표현 해상도가 향상되고 주요 주파수 성분들이 국소화 되었음을 알 수 있다. 특히 순간 주파수의 시간에 대한 기울기가 변하는 부분과 고차수 고조파 부분에서 성능 개선이 두드러짐을 알 수 있다.



(a)



(b)



(c)

그림 5. 시간-주파수 표현들의 비교

(a) Hanning 창 함수를 사용하여 얻은 스펙트로그램.

(b) 변조되지 않은 Hanning 창 함수와 제안된 방법을 사용하여 얻은 시간-주파수 표현.

(c) 변조된 Hanning 창 함수와 제안된 방법을 사용하여 얻은 시간-주파수 표현.

5. 결론 및 향후 연구

일반적으로 음성신호의 모음 부분은 다중 성분 신호이며 AM-FM 성분들로 모델링 할 수 있다. 모음 부분에서 각 AM-FM 성분들의 순간 주파수와 순간 진폭은 일정하지 않고 시간에 따라 변하는데 특히 각 성분들의 순간 주파수의 시간적 변화는 음성신호의 시간-주파수 표현의 해상도를 저하시키는 원인이 된다. 즉 시변 순간 주파수에 의한 스펙트럼의 퍼짐 현상으로 인하여 단 구간 스펙트럼에서 인접 AM-FM 성분들 간에 간섭이 일어나므로 전체 스펙트로그램의 해상도가 저하된다. 본 연구에서는 순간 주파수의 시간적 변화로 인한 스펙트로그램의 간섭현상을 최소화하여 시간-주파수 표현의 해상도를 향상시키는 방법을 제시하였다. 시간-주파수 표현을 얻기 위한 커널 함수는 주파수 변조된 함수의 형태로 되어 있어서 시변 순간 주파수를 갖는 신호를 시간-주파수 평면에 효과적으로 국소화 시킨다.

기본 순간 주파수, 즉 가장 낮은 대역의 순간 주파수를 1 차 다항식으로 근사화 하고 기울기를 추정하였으며 추정된 순간 주파수를 시간-주파수 표현을 얻기 위한 커널 함수의 FM 변조 함수로 사용하였다. 또한 창 함수를 추정된 순간 주파수에 동조 시켰다. 창 함수는 변조된 Hanning 함수를 사용하였는데 창 함수의 모양 및 길이는 고정되지 않고 순간 주파수의 기울기에 따라 변화시켜 창 효과에 의한 퍼짐 현상을 최소화 하였다. 제안된 방법을 사용할 경우 음성 신호의 시간-주파수 표

현에서 순간 주파수의 시간적 변화에 의한 해상도 저하 효과를 제거할 수 있다.

제안된 방법에서는 창 함수의 길이가 추정된 순간 주파수의 기울기의 함수인데, 순간 주파수의 기울기와는 별도로 창 함수의 길이 W를 적응적으로 변화시킬 경우 좀 더 높은 해상도를 얻을 수 있을 것으로 판단된다. 또한 순간 진폭의 시간적 변화에 의한 해상도 저하를 방지하기 위한 연구가 필요하다. 음성 신호는 다중 성분 신호인데 각 성분의 순간 주파수가 시변일 경우 기존의 방법으로는 각 성분별로 분리하는 것이 매우 어렵다. 본 연구는 음성 신호를 각 AM-FM 성분별로 분리하기 위한 사전 연구의 일환으로 수행되었다. 음성 신호의 각 AM-FM 성분별 분리를 통해 발성기관의 성문이나 성도 특성을 정확하게 파악할 수 있으며, 다양한 음색의 음성신호를 재구성하는 데 응용할 수 있다.

참 고 문 헌

[1] Cohen, L. 1995. *Time-Frequency Analysis, Englewood Cliffs*, NJ:Prentice-Hall.

[2] Rao, A. & Kumaresan, R. 2000. "On decomposing speech into modulated Components." *IEEE Tr. on Speech and Audio Processing*, Vol. 8, No. 3, 240-254.

[3] Boashash, B. 1992. *Time-frequency signal analysis*. published by Longman Cheshire.

[4] Lee, H. & and Bien, Z. 1998. "Reconstruction of signals with known instantaneous frequency using linear time-varying filter." *Electronics Letters*, Vol. 34, No. 24, 2312-2313.

[5] Qian, S. & Chen, D. 1996. *Joint time-frequency analysis: methods and applications*. published by Prentice-Hall.

[6] Griffin, D. W. & Lim, J. S. 1984. "Signal estimation from modified short-time Fourier transform." *IEEE Tr. on ASSP*, Vol. 32, No. 2, 236-243.

[7] Portnoff, M. R. 1980. "Time-frequency representation of digital signals and systems based on short-time Fourier analysis." *IEEE Tr. on ASSP*, Vol. 28, No. 1, 55-69.

[8] Lee, Heyoung & Bien, Zeungnam. 2004. "Bandpass variable-bandwidth filter for reconstruction of signals with known boundary in time-frequency domain." *IEEE Signal Processing Letters*, Vol. 11, No. 2, 160-163

[9] Quatieri, T. F., Hanna, T. E. & O'Leary, G. C. 1997. "AM-FM Separation using Auditory Motivated Filters." *IEEE Tr. on Speech Audio Processing*, Vol. 5, No. 4, 465-480.

[10] Boashash, B. 1992. "Estimating and interpreting the instantaneous frequency of a signal-part I, II," *Proc. IEEE*, Vol. 80, No. 4, 520-568.

[11] Picinbono, B. 1997. "On instantaneous amplitude and phase of signals." *IEEE Tr. on Signal Processing*, Vol. 45, No. 3, 552-560.

[12] Almeida, L. B. & Tribolet, J. M. 1983. "Nonstationary spectral modeling of voiced speech." *IEEE Tr. on ASSP*, Vol. 31, No. 6, 664-678.

[13] McAulay, R. J. & Quatieri, T. F. 1986. "Speech analysis-synthesis based on a sinusoidal representation." *IEEE Tr. on ASSP*, Vol. 34, No. 7, 744-754.

[14] Yegnanarayana, B., d'Alessandro, C. & Darsinos, V. 1998. "An iterative algorithm for decomposition of speech signals into periodic and aperiodic components." *IEEE Tr. on Speech and Audio Processing*, Vol. 6, No. 1, 1-11.

[15] d'Alessandro, C., Darsinos, V. & Yegnanarayana, B. 1998. "Effectiveness of a periodic and

aperiodic decomposition method for analysis of voice sources." *IEEE Tr. on Speech and Audio Processing*, Vol. 6, No. 1, 12-23.

[16] Cohen, L. 1989. "Time-frequency distributions: a review." *Proc. IEEE*, Vol. 77, pp. 941-981.

[17] Loughlin, P. J. & Cohen, L. 2004. "The uncertainty principle: global, local, or both?." *IEEE Tr. on Signal Processing*, Vol. 52, 1218-1227.

[18] Czerwinski, R. N. & Jones, D. L. 1997. "Adaptive short-time Fourier analysis." Signal Processing Letters, *IEEE*, Vol. 4, No. 2, 42-45.

[19] Chang, Huai Soo & Ngee Koh & Rahardja, S. 2005. "β-order MMSE spectral amplitude estimation for speech enhancement." *IEEE Tr. on Speech and Audio Processing*, Vol. 13, No. 4, 475-486.

[20] Kwok, H. K. & Jones, D. L.. 2000. "Improved instantaneous frequency estimation using an adaptive short-time Fourier transform." *IEEE Tr. on Signal Processing*, Vol. 48, No. 10, 2964-2972.

[21] Yegnanarayana, B. & Murthy, P. S. 2000. "Enhancement of reverberant speech using LP residual signal." *IEEE Tr. on Speech and Audio Processing*, Vol. 8, No. 3, 267-281.

[22] Richard, N., Czerwinski, & Douglas, L. J. 1995. "Adaptive cone-kernel time-frequency analysis." *IEEE Tr. on Signal Processing*, Vol. 43, No. 7, 1715-1719.

접수일자: 2005. 10. 31
게재결정: 2005. 11. 30

▲ 이희영
서울시 노원구 공릉 2동 172번지 (우: 139-743)
서울 산업대학교 공과대학 제어계측공학과
Tel: +82-2-970-6545 Fax: +82-2-949-2654
E-mail: leehy@snut.ac.kr

▲ 최승호
서울시 노원구 공릉 2동 172번지 (우: 139-743)
서울 산업대학교 공과대학 전자정보공학과
Tel: +82-2-970-6461 Fax: +82-2-979-7903
E-mail: shchoi@snut.ac.kr