pISSN 2005-8063
eISSN 2586-5854
2018. 12. 31.
Vol.10 No.4
pp. 121-125

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

https://doi.org/10.13064/KSSS.2018.10.4.121

Jitter 합성에 의한 음질변환에 관한 연구

Voice quality transform using jitter synthesis

조 철 우*

Jo, Cheolwoo

Abstract

This paper describes procedures of changing and measuring voice quality in terms of jitter. Jitter synthesis method was applied to the TD-PSOLA analysis system of the Praat software. The jitter component is synthesized based on a Gaussian random noise model. The TD-PSOLA re-synthesize process is used to synthesize the modified voice with artificial jitter. Various vocal jitter parameters are used to measure the change in quality caused by artificial systematic jitter change. Synthetic vowels, natural vowels and short sentences are used to check the change in voice quality through the synthesizer model. The results shows that the suggested method is useful for voice quality control in a limited way and can be used to alter the jitter component of voice.

Keywords: jitter, TD-PSOLA, resynthesis, voice quality

1. 서론

음성의 음원의 변화를 나타내는 불규칙적인 변수 중에는 진폭 성분의 변화와 주기성분의 변화가 있다. 이러한 불규칙적 변화는 음성의 자연성과 연관되어 있다고 생각되나 피치주기 등과는 달리 상대적으로 널리 연구의 대상이 되지 않고 있다. 이러한 주기성분의 불규칙성 때문에 음성은 준 주기성을 갖는다고도 한다. 1968년 Fujimura에 의해 스펙트럼 대역에서의 준주기적 현상이 연구된 이래(Fujimura, 1968) Titze, Deshmukh 등이 음성의 준주기성 검출에 관하여 연구하고 보고한 바 있다(Titze, 1995; Deshmukh *et al*., 2005). 음성합성에의 적용 사례로는 Hillenbrand가 합성음에서의 주기성의 변이에 대해 보고한 바가 있고(Hillenbrand, 1987) Endo & Kasuya 등이 장애음성의 합성에 적용한 사례가 있다(Endo & Kasuya, 1996). 국내에서는 음성의 준주기성인 jitter의 특성에 관해 조사하고 가우시안 분포를 통하여 합성해 내는 방법에 관한 연구가 진행된 바가 있다. 이 논문에서는 선행 연구자들의 방법을 통해 jitter 성분을 체계적으로 제어하는 방법을 합성음성을 통하여 실험하고 실제 발성음성에서 측정한 jitter 성분의 범위와 유사한 형태로 생성될 수 있도록 하여 방법의 유효성을 확인한 바 있다(조철우, 2016). 그러나 선행 연구에서는 jitter 성분을 제어하는 방법을 실제 음성에 대해 적용하고 그 효과를 측정한 사례가 부족하였으므로 본 연구에서는 이를 보완하여 실제 음성을 음성합성 모델을 통해 분석하고 분석된 주기성분을 변형하여 재합성한 뒤 다양한 음

* 창원대학교, cwjo@changwon.ac.kr

Received 1 November 2018; Revised 10 December 2018; Accepted 10 December 2018

성의 준 주기성과 관련한 파라미터들의 측정을 통해 인위적으로 변화시킨 준주기성 변화에 따른 음질의 변화를 조사하고자 한다.

2. Jitter 합성방법

준주기성의 구현 방법은 먼저 특정 F_0의 주기에 대해 일정한 변이된 주기를 1개 혹은 여러 개의 간격으로 삽입하는 방법(Kreiman & Gerratt, 2003), 두 번째는 하모닉 상태의 음원을 생성한 뒤 특정 전달함수를 통과시켜 F_0의 변이를 생성하는 방법(Endo & Kasuya, 1996) 그리고 세 번째로 특정 F_0을 기준으로 난수발생기에 따른 F_0의 변이를 생성하여 구현하는 방법이 있다(Alzamendi *et al*., 2013; Ruinskiy & Lavner, 2008). 앞의 두 가지 방법은 정확한 jitter값을 구현하는 데는 적합하지 않기 때문에 세 번째 방법을 사용하였다. 먼저 정해진 jitter값을 갖는 준주기적 펄스열을 생성하는 데 필요한 확률분포를 측정하고 이를 바탕으로 통계모델을 구한 다음 그 결과로부터 난수발생기를 통해 펄스열을 발생시킨다.

유성음의 피치 구간분석 결과로부터 우리는 피치주기의 변동이 정규분포의 형태와 유사하다는 것을 확인할 수 있었다. 이에 따라 피치 주기의 변이를 정규분포에 따라 발생시킨다는 가정 하에 원하는 jitter값을 F_0의 변이로부터 구하는 식은 다음과 같이 도출된다. Jitter%값은 다음과 같이 정의된다고 한다.

$$jitter\% = 100\frac{\frac{1}{N-1}\sum_{j=1}^{N-1}|P_{j+1}-P_j|}{\frac{1}{N}\sum_{j=1}^{N}P_j} \tag{1}$$

준주기적 펄스의 변동이 가우시안 분포를 따른다고 가정하였을 때 확률 밀도함수를 다음과 같이 정의할 수 있다. $N(P_0,\sigma)$. 여기서 P_0는 주기의 평균값이고 σ는 표준편차이다. 이 가정은 Titze(1995)와 Kreiman(Kreiman & Gerratt, 2003)에 의해 사용되었다. 밀도함수에 기반하여 각 피치 위치의 변동의 폭이 결정된다. 음성신호의 주기의 변이 분포는 F_0의 역수에 해당하는 주기의 빈도가 크고 주기가 짧아질수록 빈도가 줄어드는 현상을 보이는데 이는 가우시안 분포의 절반에 해당하는 반가우시안 분포의 특성을 갖는다고 가정할 수 있다.

$\triangle P_j = P_{j+1} - P_j$ 여기서 $j = 1,...,N-1$이다.

$|\triangle P_j|$는 반가우시안(hemi-Gaussian) 특성을 가지며 다음과 같은 확률밀도함수를 갖는다.

$$\begin{cases} N(0,\sqrt{2}\,\sigma_P)\ , & \text{if } |\triangle P_j| = 0; \\ 2N(0,\sqrt{2}\,\sigma_P), & \text{if } |\triangle P_j| > 0; \\ 0\text{, 그 외에.} \end{cases} \tag{2}$$

$$E\{|\triangle P_j|\} = \int_0^{\infty} \frac{2|\triangle P_j|\exp\left(\frac{-|\triangle P_j|^2}{4\sigma_P^2}\right)}{(4\pi\sigma_P^2)^{1/2}} d|\triangle P_j| = \frac{2\sigma_P}{\sqrt{\pi}} \tag{3}$$

$N\to\infty$이면 $\left\{\frac{1}{N-1}\sum_{j=1}^{N-1}|P_{j+1}-P_j|\right\}$는 $E\{|\triangle P_j|\}$에 수렴하며 $\left\{\frac{1}{N}\sum_{j=1}^{N}P_j\right\}$는 P_0에 수렴한다는 것이 알려져 있다. 식 (3)을 식 (2)에 대입하여 풀면

$$\sigma_P = \frac{\sqrt{\pi}\,P_0 jitter\%}{200} \tag{4}$$

와 같이 구해진다(Alzamendi *et al*., 2013). 식 (4)에 의해 원하는 jitter값을 갖는 펄스위치의 변이를 만들기 위한 위치의 편차를 동일한 분포를 갖는 난수발생기를 이용하여 구할 수 있다.

주기성분은 주기적인 펄스 형태로 모델링할 수 있다. 준주기적 현상은 정확히 주기적인 펄스의 주기를 어떤 방식으로 임의적인 변화를 주도록 함으로써 구현할 수 있다. 준주기성을 구현하는 방법은 두 가지가 있다. 첫 번째는 주기적인 펄스를 통해 주기성분을 구현한 뒤 펄스의 위치를 랜덤하게 변경시켜줌으로써 구현하는 방법이 있고(Endo & Kasuya, 1996), 두 번째는 주기적인 펄스를 생성한 뒤 난수 성분이 가미된 준주기성 펄스를 만드는 방법이 있다(Hillenbrand, 1987). 본 연구에서는 첫 번째 방법을 사용하였다. 실험에서는 특정 jitter값을 가정하지 않고 임의로 순서적으로 값을 증가시켜가며 발생시키는 방식을 적용하였다.

Jitter 성분 생성을 위해서 가우시안 백색 잡음을 발생시키기 위하여 Praat에서 제공하는 randomGauss함수를 사용한다. 그림 1은 함수에 의해 생성된 50개의 잡음 신호 값의 히스토그램 분포를 보여준다.

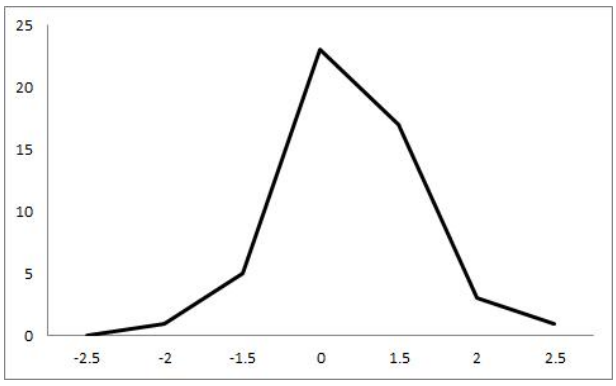

그림 1. RandomGauss 함수로부터 발생한 난수의 분포
Figure 1. Histogram of random numbers from the function randomGauss

3. 합성음성의 음질변환

3.1. 합성음 생성 및 변환

Praat코드로부터 jitter값 변경에 대한 효과를 측정하기 위하여 먼저 파라미터를 제어하기 쉬운 합성음성을 합성하고 합성된 음성에 피치값을 체계적으로 변동시켜가며 음질의 변화를 측정하고자 하였다. 합성음성의 생성은 Praat에 내장된 KlattGrid 음성합성 모델을 이용하여 음성을 합성하였다. KlattGrid는 기존의 D. H. Klatt의 합성기를 토대로 음원 등에 일부 변경한 구조를 가지고 있다. 합성은 제1, 제2, 제3포먼트 만에 의해 모음을 합성하였다. 피치 궤적은 단조롭게 감소하는 형태로 합성하였다. 합성한 음성은 TD-PSOLA 방법에 의해 분석하여 jitter값을 반영하고 결과음성을 TD-PSOLA 모델에 의해 다시 합성하였다. 그림 2는 KlattGrid 음성합성 방법의 원리를 나타낸다. 분석에 TD-PSOLA 방식을 적용한 이유는 차후 자연발화음성에 jitter 변동을 구현하고 측정할 때 이 방식을 사용하므로 방법의 일관성을 유지하기 위함이다. TD-PSOLA 방식은 각 음성의 음원의 박동기를 검출하고 이를 바탕으로 음원진동 주기를 검출한 뒤 이 신호를 통해서 음원의 주기성 진동을 제어하게 된다. Praat의 TD-PSOLA 합성모델에서는 Point process에 의해 음원 박동기점이 효과적으로 분리되게 된다. Jitter 생성을 위해서는 TD-PSOLA 합성기의 pitch tier를 구하고 여기에 잡음성분을 더하여 피치구간의 변동량을 임의적으로 변화하도록 하여 Voicing 블럭을 변경시켜주면 된다. 그림 3에 전체 변환 과정을 나타내었다.

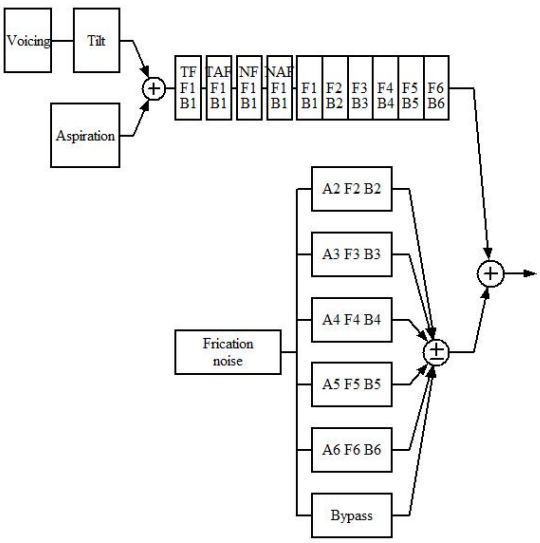

그림 2. KlattGrid 합성기 구조
Figure 2. Structure of KlattGrid synthesizer

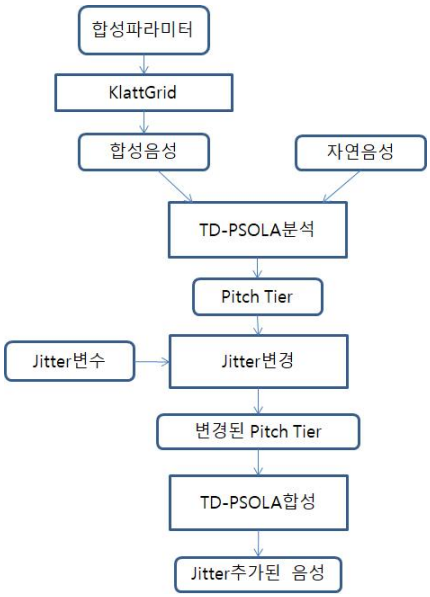

그림 3. 음성변환과정
Figure 3. Voice conversion process

3.2. 자연발화음성의 음질변환

녹음된 자연발화음성에 합성음성과 동일한 방법으로 jitter값을 변화시켰을 경우 음질의 변화에 어떻게 영향을 미치는지 관찰하기 위해 음성을 녹음하고 jitter값을 변경시킨 이후 음질을 측정하였다. 자연발화음성은 길이가 짧은 문장음성(“안녕하십니까?”, 남성화자)을 사용하여 모음부분의 중심부에서 파라미터를 측정하였다.

자연발화 음성으로부터 jitter 성분이 추가된 합성음성의 생성은 praat에 내장된 분석법에 의해 pitch tier를 구한 뒤 이 값들에 랜덤함수값을 통해 각 피치 주기의 시작점을 변동시킨 후 재 합성하는 방법을 사용하였다. Jitter 성분의 구현은 합성음성의 경우와 동일한 방법을 적용하여 랜덤 함수의 분산값이 서로 다른 10 단계로 증가하는 jitter 변수에 대하여 Praat에 내장된 TD-PSOLA 음성합성 모델을 이용하여 음성을 분석하고 다시 합성하였다.

4. 음질 측정 및 평가

Jitter 성분을 변화시켜 재합성한 음성의 음질을 측정하기 위하여 Praat에 내장된 jitter파라미터들과 피치 변수에 따른 파라미터들을 측정하여 음질의 변화를 관찰하였다. 측정한 파라미터들은 다음과 같다. Jitter(local), RAP(relative average perturbation), PPQ5(five point period perturbation quotient), DDP(difference of differences of periods). 이들 파라미터는 피치변동량을 조금씩 다른 관점에서 나타내는 파라미터로 준주기성 관련한 음질을 정량적으로 나타내는 수치들이다. 각 연관된 파라미터의 의미는 다음과 같다. Jitter는 연속한 피치구간에서의 피치값의 평균 변동량을 나타낸다. RAP는 인접한 두 개의 이웃한 구간에서의 피치변동값의 상대적 변화를 나타낸다. PPQ5는 인접한 4개의 분

석구간에서의 jitter값의 변동을 나타낸다. DDP는 인접한 jitter값의 2차미분에 해당한다. 이들 파라미터는 피치구간의 불규칙성을 서로 다른 척도로 표시해주고 있으므로 피치 변동에 의한 음질의 변화를 측정하는 파라미터로 널리 사용되고 있다. 본 연구에서는 오로지 피치 변동과 결과로 얻어지는 주기성의 변동의 연관성에 초점을 두었으므로 주관적인 음질평가에의 영향은 배제하기로 한다.

파라미터 변동 값의 범위를 설정하기 위하여 랜덤함수의 분산값을 1부터 10까지로 단계별로 변화시키면서 변환된 음성으로부터 jitter 및 음성파라미터들을 구하였다.

매 실행시 나타나는 경향은 분산값의 증가에 따라 일관성 있게 증가하지 않고 랜덤하게 증가하는 현상을 보였다. 3회 반복 수행한 후 회귀직선을 구하고 이를 통하여 jitter값의 변화 경향을 추정하였다.

그림 4는 합성음성의 경우에 분산값을 1부터 10까지 변화시켰을 경우 나타난 jitter값의 변화 경향을 보여준다. 측정값의 분산도가 커서 뚜렷한 경향을 볼 수는 없었으나 다중 측정 및 회귀직선을 통하여 일정한 기울기로 증가하는 경향을 관찰할 수 있었다. 그림 5는 자연발화 음성을 변경하여 재합성한 음성으로부터 측정한 jitter값의 변화 경향을 보여준다. 합성음성의 경우 자연발화 음성보다 약간 더 큰 jitter값의 변화를 보여주고 있는 것이 관찰된다. 그림 6은 RAP, 그림 7은 PPQ5, 그림 8은 DDP값의 변화경향을 보여준다. Jitter와 마찬가지로 전체적으로 값이 증가하는 패턴을 보이고 있다.

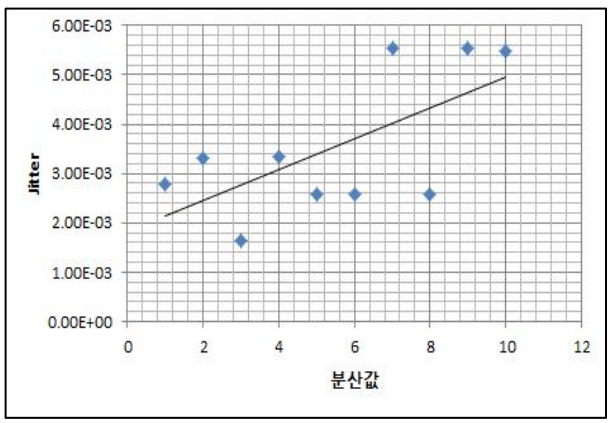

그림 4. 변환된 합성음성으로부터 jitter 측정결과
Figure 4. Measurement of jitter from transformed synthetic voice

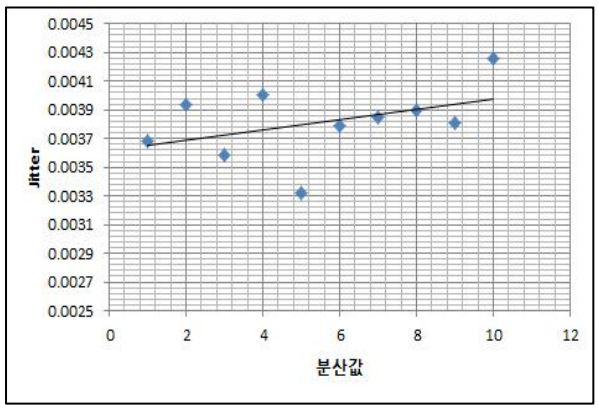

그림 5. 변환된 자연음성으로부터 jitter 측정결과
Figure 5. Measurement of jitter from transformed natural voice

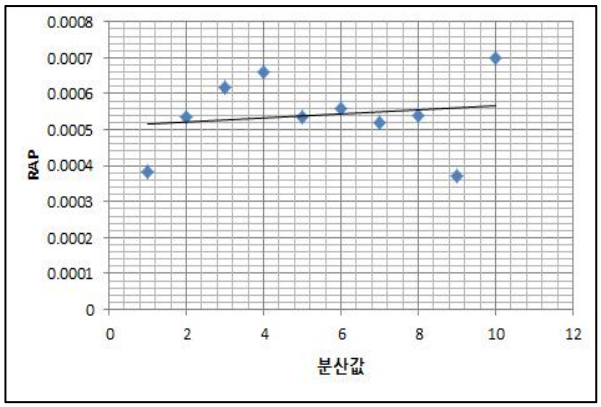

그림 6. 변환된 자연음성으로부터 RAP 측정결과
Figure 6. Measurement of RAP from transformed natural voice

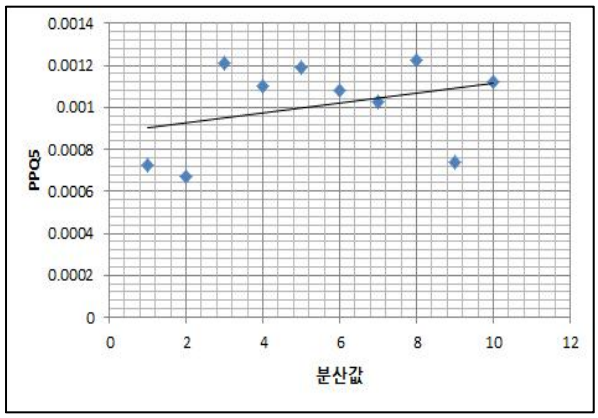

그림 7. 변환된 자연음성으로부터 PPQ5 측정결과
Figure 7. Measurement of PPQ5 from transformed natural voice

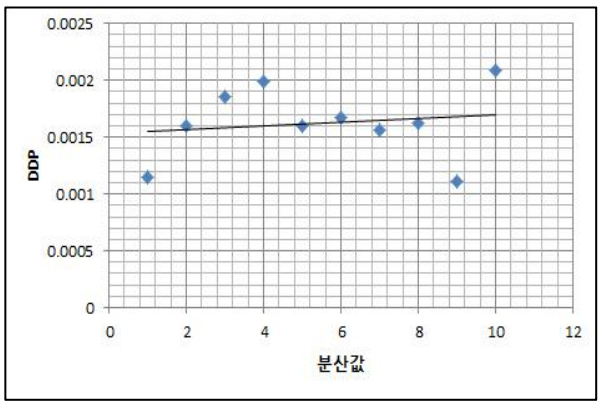

그림 8. 변환된 자연음성으로부터 DDP 측정결과
Figure 8. Measurement of DDP from transformed natural voice

5. 결론

본 연구에서는 합성된 음성과 자연음성의 피치성분을 랜덤함수를 통해 변경하여 재합성하면서 이에 따른 jitter 성분 및 이와 연관된 파라미터들을 측정하여 그 변화의 경향을 살펴보았다. Jitter 성분이 발생하는 원리는 음원의 진동의 불규칙성이라는 데 근거하여 음성합성기를 통한 음원모델로부터 주기성에 랜덤함수를 통하여 불규칙성을 가하여 실험을 했다. 실험결과 주기변동의 랜덤 값의 크기에 비례하여 증가하는 현상을 확인하였다.

그러나 변동폭이 과도하게 커질 경우 자연발화 음성에 나타나는 범위를 초과하여 변동하여 부자연스럽게 되는 현상이 있었으며 일정한 구간에 한해서 jitter 성분의 제어가 가능하다는 것을 확인하였다. 실험에 사용한 음성합성 모델은 실제 사람의 음성모델과는 차이가 있으므로 일반적인 음성의 생성 현상에 적용하는 데는 무리가 있으나 장애음성 합성 등 시뮬레이션을 통한 응용에 적용할 수 있는 경향을 확인했음에 의의가 있다고 본다. 향후 jitter 성분을 보다 정교하게 가변할 수 있는 모델의 개발이 필요하다.

감사의 글

이 논문은 2017–2018년도 창원대학교 자율연구과제 연구비 지원으로 수행된 연구결과입니다.

참고문헌

Alzamendi, G. A., Schlotthauer, G., Rufiner, H. L., & Torres, M. E. (2013). Evaluation of a new model for vowels synthesis with perturbations in acoustic parameters. *Latin American Applied Research*, 43(3), 225-230.

Deshmukh, O., Espy-Wilson, C. Y., Salomon, A., & Singh, J. (2005). Use of temporal information: Detection of periodicity, aperiodicity, and pitch in speech. *IEEE Transactions on Speech and Audio Processing*, 13(5), 776-786.

Endo, Y., & Kasuya, H. (1996). A stochastic model of fundamental period perturbation and its application to perception of pathological voice quality. *Proceedings of ICSLP'96*, Philadelpia, PA, USA. October 3-6, 1996.

Fujimura, O. (1968). An approximation to voice aperiodicity. *IEEE Transactions on Audio and Electroacoustics*, 16(1), 68-72.

Hillenbrand, J. (1987). A methodological study of perturbation and additive noise in synthetically generated voice signals, *Journal of Speech*, *Language*, *and Hearing Research*, 30(4), 448-461.

Jo, C. (2016). Analysis and synthesis of pseudo-periodicity on voice using source model approach. *Phonetics and Speech Sciences*, 8(4), 89-95. (조철우 (2016). 음성의 준주기적 현상 분석 및 구현에 관한 연구, *말소리와 음성과학*, 8(4), 89-95.)

Kreiman, J., & Gerratt, B. R. (2003). Jitter, shimmer, and noise in pathological voice quality perception. *Proceedings of VOQUAL 2003* (pp. 57-61). Geneva, Switzerland.

Ruinskiy, D., & Lavner, Y. (2008). Stochastic models of pitch jitter and amplitude shimmer for voice modification. *Proceedings of 2008 IEEE 25th Convention of Electrical and Electronics Engineers in Israel* (pp. 489-493). Eliat, Israel. December 3-5, 2008.

Titze, I. R. (1995). *Workshop on acoustic voice analysis: Summary statement*. Denver: National Center for Voice and Speech.

• **조철우 (Jo, Cheolwoo)**
창원대학교 전기전자제어공학부 교수
경남 창원시 의창구 창원대학로 1
Tel: 055-213-3662 Fax: 055-262-5064
Email: cwjo@changwon.ac.kr
관심분야: 음성신호처리, 장애음성분석