



Statistical analysis on long-term change of jitter component on continuous speech signal*

Cheolwoo Jo**

School of Electrical, Electronics and Control Engineering, Changwon National University, Changwon, Korea

Abstract

In this study, a method for measuring the jitter component in continuous speech is presented. In the conventional jitter measurement method, pitch variabilities are commonly measured from the sustained vowels. In the case of continuous speech, such as a spoken sentence, distortion occurs with the existing measurement method owing to the influence of prosody information according to the sentence. Therefore, we propose a method to reduce the pitch fluctuations of prosody information in continuous speech. To remove this pitch fluctuation component, a curve representing the fluctuation is obtained via polynomial interpolation for the pitch track in the analysis interval, and the shift is removed according to the curve. Subsequently, the variability of the pitch frequency is obtained by a method of measuring jitter from the trajectory of the pitch from which the shift is removed. To measure the effects of the proposed method, parameter values before and after the operations are compared using samples from the Kay Pentax MEEI database. The statistical analysis of the experimental results showed that jitter components from the continuous speech can be measured effectively by proposed method and the values are comparable to the parameters of sustained vowel from the same speaker.

Keywords: Jitter, connected speech, variability measure, polynomial interpolation, sustained vowel

1. 서론

음성신호의 불규칙성은 생체기관의 불규칙한 운동에 기인하는 것이다. 음성신호의 음원 분석에 있어서 기본적인 가정은 성대의 진동이 준주기적이라는 것이다. 이 과정에서 사실상 주기성을 검출하는 자기상관법이나 주파수성분에서의 하모닉 성분을 검출하는 방법을 사용해 왔다. 그러나 실제 음성은 완전한 주기성을 갖지 않으므로 주기성을 검출하는 분석 과정에서 음

성신호의 자연성을 구현하고 있는 불규칙성에 대한 정보가 손실되게 된다. 주기 성분 분석에 있어서 분석구간으로 잡고 있는 20ms 정도의 구간은 구간 내에서 발성기관의 변동성이 미미하여 동일한 진동을 생성한다고 가정한 것에 근거를 두고 있다. 따라서 통상적인 주기성 분석에 의한 결과는 기본적으로 20ms 구간을 기본으로 한 저역통과 필터된 값이라고 볼 수 있다. Manfredi et al.(2011)에 따르면 합성된 신호를 분석하여 검증한 결과 동일한 음성 시료에 대해서도 분석 도구에 따라 구해진

* This research was supported by Changwon National University in 2019~2020.

** cwjo@changwon.ac.kr, Corresponding author

Received 18 October 2020; Revised 17 December 2020; Accepted 17 December 2020

© Copyright 2020 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

jitter값의 오차가 달라진다는 것이 알려졌다. 그러므로 일관된 값을 제공하는 또 다른 분석 방법이 필요하다. Fourcin(2009)은 연속 음성에서 음성의 불규칙성을 측정하기 위하여 생체신호로 직접 측정된 성대 박동신호(EGG)를 사용하여 시간에 따른 주기성의 변동을 측정하였다. 이 경우에는 EGG를 측정하기 위한 특별한 장치를 통해 신호를 수집하고 분석하였다. 이 문헌의 결론에서 저자는 생체신호를 활용하지 않고 음향 신호만에 의해서도 충분히 연속적인 불규칙 현상을 측정할 수 있다고 결론을 내리고 있다. Kisenwether & Prosek(2017)는 음성의 주기성과 진폭에서의 불규칙성을 측정하기 위하여 단구간 분석과 장시간 분석을 모두 수행하여야 하는 가에 유의하여 기존의 불규칙 파라미터 분석 방법을 적용하여 단구간 분석과 장시간 분석의 결과를 고찰하고 비교하였다. 결론적으로 임상적 목적의 분석에서 단구간 분석과 장시간 분석의 결과의 차이가 미미하여 단구간 분석으로도 장애음성 판별등의 목적에 부합하는 결과를 얻을 수 있다고 결론을 내리고 있다. 그러나 이 연구에서는 MDVP 데이터에 포함된 지속모음에 한정하여 분석하였고 동일한 DB에 포함된 문장 발성을 분석하지 않았으므로 연구의 범위가 제한적이었다. 모음의 연속 발성과 단구간 분석에서의 차이점을 고찰한 Gerratt et al.(2016)은 지속모음과 문장음성에서의 분석 결과의 차이를 보이기는 하나 안정구간의 모음에 한정되어 분석하였고 전체적인 운율 변동을 고려하지는 못하였다. 연구에서 주 관심이었던 음원의 변화보다는 운율적인 내용의 변화에 주로 기인하는 것으로 결론을 내리고 있다. Dajer (2005)는 음성 특성의 동적 변화를 관찰하기 위해 비선형 분석 기법을 적용하기도 하였으나 지속모음에 한정하였다.

Schoentgen(1989)은 Dysphonic Speaker에 대한 실험에서 고립된 문장에서 분리한 /a/, /e/, /i/모음의 안정된 구간에서 측정된 Jitter값을 통한 정상군과 대비한 변별력이 지속모음에서 측정된 파라미터 값에 비해 낮지 않다고 결론짓고 있다. 이 방법에서 고립된 문장에 대해 사용한 jitter값 측정 방법은 PPQ(Period Perturbation Quotient)로 구간별 평균값을 영향을 고려하였다. 그러나 인접 5구간의 평균값만을 고려하여 전체 운율 정보의 영향을 측정하는 데는 한계가 있다.

또한 Schoentgen & Guchteneere(1995)는 지속모음에 대해 시계열분석 방법에 의한 jitter측정을 시도하였다. 이를 통해 화자간 변동성을 축소할 수 있었다고 보고하고 있다. Vasilakis & Stylianou(2009)는 Running speech에 대한 jitter추정 방법으로 스펙트럼 상에서 하모닉 성분을 통한 jitter추정 방법인 SJE(Spectral Jitter Estimator)를 제안하여 사용하고 있다. 이 방법에서는 제안한 SJE방법을 문장음성에 지속적으로 적용한 뒤 정해진 역치값을 통해 시간축에서 부분적으로 정상음성과 장애음성을 식별하는 방법을 제안하였다. 이 방법에도 4피치 구간 길이를 통해 스펙트럼을 구하는 방식을 적용하였으므로 운율 정보의 영향 감소는 PPQ와 비슷한 수준으로 판단되며 단지 잡음 환경에서 좀 더 강인한 특성을 갖는다.

기존의 연구 사례를 살펴볼 때 지속 모음에 의한 기존의 Jitter 산출 방식은 단순하고 효과적인 측정법이기는 하나 동시에 지

속 모음 발성은 문장음성 발성 또는 대화음성과 같은 자연스러운 발성상태와는 차이가 있기 때문에 자연스러운 연속 발화로 부터 피치 값의 변동성을 측정하려는 시도가 지속적으로 있어 왔다. 또한 문장음성의 경우 화자의 조음상태의 변화도 포함하고 있으므로 성대의 떨림 외에 다양한 정보를 포함하고 있어서 다양한 경우에 적용할 가능성을 갖고 있다. 지금까지의 연구에서는 전체적인 운율의 영향을 제거하기 보다는 안정된 구간에 대해 인접 피치주기의 평균을 고려하는 형태로 문장음성내의 피치 변동성을 측정해 왔다. 문장음성을 통한 연구의 사례에서는 접근 방법에 따라 서로 다른 음성군의 식별에 있어서는 지속 모음의 경우보다 못한 경우도 있는 반면 방법과 시료에 따라서 더 나은 결과를 얻는 사례도 있으므로 문장음성을 통한 피치의 변동성 측정을 시도해볼만한 가능성이 여전히 존재한다.

본 연구에서는 음원의 불규칙성 측정의 관점에서 지속모음이 아닌 문장 발성에서의 음성 신호로부터 화자 발성의 주기성에서의 불규칙성을 측정하는 방법과 그 결과에 관하여 고찰하고자 한다.

2. 피치 값 변동의 측정 방법

Jitter값은 단 구간 내에서 피치 값의 변동성을 나타내는 척도로 음원의 변동을 잘 반영해 주는 파라미터로 널리 활용되어 왔다. 기존의 주기적 불규칙성을 측정하는 방법인 jitter값이 문장 내에서 어떻게 변화하는지를 측정하여 기본 분석 값으로 삼고자 한다. Jitter는 주기측정 값인 피치 값을 단구간 분석에 의거하고 그 구간내에서 피치값의 변동율을 구하는 것이다. 문장 음성에서 유성음 구간을 추출한 뒤 각 구간의 jitter값 및 전체 jitter값의 통계적 분포를 단구간 발성에서 구한 jitter값과 비교하여 값의 변동을 고찰한다.

2.1. Jitter

음성신호의 피치 값의 변동을 나타내는 jitter는 주기 변동의 평균값을 나타내는 Jita가 있고 연관 파라미터로 주기변동의 평균 주기와의 비율을 나타내는 Jitt, 인접한 3 주기간의 변동폭을 고려한 RAP(Relative Average Perturbation), 인접한 5주기간의 변동폭을 고려한 PPQ, 주기 값의 2차 차분인 DDP(Differences of Differences of Period)등이 널리 사용되고 있다. 이들 값은 기본적으로 주어진 구간 내에서 평균적인 피치값의 변화를 측정한다. Jita는 다음과 같이 정의된다.

$$Jita = \frac{1}{N} \sum_{j=1}^{N-1} |P_{j+1} - P_j| \quad (1)$$

P_j 는 j번째 피치값을, N은 구간내의 전체 피치 값의 개수를 의미한다.

2.2. 문장 내에서의 jitter 의 변동값 측정

동일 화자의 경우라도 문장 내에서의 jitter값은 변한다. 모음의 종류나 발성 방법에 따라라도 변화할 수 있다. 문장 내에서의 jitter값은 문장의 피치값 변동에 따라 영향을 받을 수 있으나 PPQ 또는 DDP 파라미터 값은 인접 피치값과의 평균값을 통해 구간내 피치값 변동의 영향을 줄여주므로 일종의 저역통과 필터의 효과가 있다. 문장음성의 경우 다양한 모음이 존재한다. 여러 모음에 따른 jitter값의 변화에 관하여 MacCallum et al.(2011)은 모음 /a/에 비해 /i/나 /u/가 높은 피치값을 갖는 동시에 더 큰 신호대 잡음비를 갖는다고 비선형 분석을 통해 결론을 내렸으나 jitter나 shimmer는 모음간의 음향적 차이를 구분하는데 적절하지 않다고 하였다. 이 결론은 바꾸어 말하면 모음의 종류에 따른 음향적 특성의 차이가 구분되기 어렵다는 것이다. Gerrat et al.(2016)은 문장 내에서의 하모닉 성분을 통한 음질분석에서 문장 내에서 3-4개의 서로 다른 모음의 안정된 구간으로부터 얻은 데이터를 분석하고 있다. 이 경우는 하모닉 성분 추출을 스펙트럼분석과 켈스트럼 분석에 의해 수행하였으므로 분석 방법의 특성상 일정 구간 이상의 길이의 표본을 필요로 하였다. 이 경우에 피치 구간의 변동이 스펙트럼을 구하는 과정에 흡수되게 되므로 구간내에서의 피치 값의 변동은 평균값이 되어 파라미터 측정에 영향을 미칠 수 없게 된다. 그러나 이 실험에서 이용한 방법은 안정된 모음구간만을 따로 추출하였기 때문에 운율정보의 변화에 따른 피치 트랙의 변화를 고려하지는 않았다. Vasilakis & Stylianou(2009)은 running speech에서 피치 변동의 영향을 축소하기 위하여 단구간 스펙트럼에 의한 하모닉스 측정 방법 또는 전체 유성음 구간에 대해 구한 피치 값들을 통해 jitter값을 구한 뒤 통계적으로 구한 jitter의 역치값과 구간별 값의 초과 여부를 측정하여 정상음성과 장애음성의 식별율을 단모음에 의한 결과와 비교하였다. Teixeira & Golcalves(2014)에 의하면 신호 합성에 의해 동일한 피치값으로 측정된 Jitter파라미터 값도 피치값의 변이가 어떤 형태로 가해지는가에 따라 서로 다른 jitter값이 측정된다고 결론지었다. 그러므로 문장음성 또는 연속음성과 같이 지속적으로 피치값이 변화하는 경우는 jitter파라미터의 측정은 지속모음을 측정하는 방법과는 다른 방법으로 측정되어야 하며 측정 값에 대한 판단기준도 달라야 한다.

본 연구에서는 이러한 선행 연구들에 기반하여 문장 내에서 다양한 모음이 존재하는 rainbow speech의 경우 모음 구간을 추출한 뒤에 구간별로 피치의 박동기를 구하고 이를 바탕으로 jitter값을 구할 경우 화자의 음원의 변동성을 일관성있게 나타낼 수 있다고 보고 문장단위의 피치 변동성 분석을 축소하는 기법을 제안하고 수행하였다.

연속 음성에서의 피치 변동은 다음과 같은 특성을 갖는다.

1. 유성음의 구간 길이가 가변적이다.
2. 분석 구간내에서 피치값의 변동이 있다.
3. 때로는 분석 구간에서의 급격한 변동이 있다.

피치 세그먼트내에서의 분석 순서는 다음과 같다.

1. Praat를 이용하여 음성신호의 피치의 위치를 추출하는 PointProcess분석을 행한다. 이 과정은 자기상관법에 의해 피치 박동기를 구하는 방법이다. 이 과정에서 유성음 구간의 추출도 같이 행해진다.
2. PointProcess결과로부터 위치 정보를 읽어들인다.
3. 위치정보로부터 문장 전체의 피치값 벡터를 구성한다.
4. 유성음 구간별 피치값의 트렌드를 다항식 보간법에 의해 제거한다.
5. 변동성이 제거된 변동 곡선을 결정한다.
6. 구해진 변동곡선을 원래 트랙으로부터 차감하여 트랙을 평탄화 한다.
7. 평탄화된 트랙으로부터 jitter 정보를 구한다.
8. 모든 피치 세그먼트에 대해 반복 분석한다.

2.3. 피치 세그먼트 jitter 모델

연속적인 음성신호는 유성음, 무성음, 묵음 구간으로 나뉜다. 유성음 구간에는 성대의 진동에 기반한 피치가 주기적으로 나타난다. 연속 음성 내에서 피치값이 연속적으로 나타나서 유성음 구간이 끝나는 곳까지를 피치 세그먼트라고 하자. Jitter 성분의 측정은 하나의 피치 세그먼트 내에서 하게된다.

하나의 피치 세그먼트 내에서의 피치 변동은 성대의 떨림의 불규칙적 운동에 의해 영향을 받아 피치 트랙의 변이가 더해진다. 그림 1은 세그먼트 내의 jitter생성 모델이다.

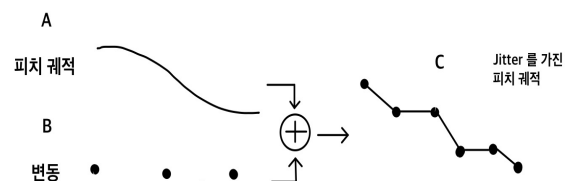


그림 1. 피치세그먼트 Jitter생성 모델
Figure 1. Jitter production model of pitch segment

만약 피치 궤적의 형태를 적절히 추정할 수 있다면 $B=C-A$ 와 같이 연산에 의해 음원의 변동을 근사하게 추정할 수 있다.

유성음 구간에서 피치정보의 트렌드 제거는 다음과 같이 한다. 분석 구간 내에서 피치 정보를 평탄화 하기 위하여 기준 값으로 다항식 보간법에 의한 경향을 구한다. 다항식 보간은 다음과 같이 한다.

구간내의 피치 궤적을 y 라고 할 때 y 를 근사하는 다항식 곡선의 방정식을 구한다. 보간식을 구할 때 전체적인 경향만 포함하도록 다항식의 차수를 포함한 표본의 개수보다 적게 선택하면 피치값의 변이를 제거한 성분만을 추정할 수 있다.

보간 다항식은 다음과 같이 설정한다.

만약 n 개의 데이터가 주어졌다고 가정하자.

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n) \quad (2)$$

이 n개의 데이터를 가장 잘 표현하는 다항식을 구하는 것이 목표이다. 다항식은 다음과 같이 m차의 다항식으로 표현할 수 있다.

$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_mx^m \quad (3)$$

여기서 주어진 x, y 값들을 가장 잘 표현하는 계수 $a_0, a_1, a_2, \dots, a_m$ 을 구하는 것이다.

주어진 계수로 표현한 다항식의 값과 원래 값의 오차는 다음과 같이 표현할 수 있다.

$$E_i = y_i - (a_0 + a_1x_i + a_2x_i^2 + \dots + a_mx_i^m) \quad (4)$$

$$\frac{\delta}{\delta a_j} \sum_{i=1}^m |E_i|^2 = 0 \quad (j = 1, 2, 3, \dots, m) \quad (5)$$

이 되는 $a_0, a_1, a_2, \dots, a_m$ 를 구하면 주어진 값에 대응하는 다항식을 구할 수 있다.

이때 차수 m을 부족 조건이 되도록 설정하면 전체적인 데이터의 전체적인 동향은 유지하되 세부적인 변동을 제거한 형태의 궤적을 얻을 수 있다. 보통의 다항식 피팅에서는 주어진 데이터를 충실히 반영하는 것을 목표로 하고 있으나 주어진 데이터의 개수에 비해 다항식의 차수를 현저히 낮게 한다면 변동 경향만을 나타내는 곡선의 식을 구할 수 있다.

부족 보간 조건은 분석 구간에 포함되는 피치값의 개수를 참조하여 경험적으로 정한다. 일반적으로 다항식 보간법에서 적정 보간의 차수를 정하는 방법은 표본의 개수보다 큰 값으로 차수를 정하는 것이 일반적이나 여기서는 전체적인 경향만을 대표하는 곡선을 구하기 위하여 부족 보간이 되도록 포함된 표본의 개수보다 훨씬 적은 값을 차수로 택하도록 하였다. 이 실험에서는 경험에 의한 선택을 통해 4-7사이의 차수가 적절하였고 그 이상의 차수를 적용하였을 경우 곡선의 굴곡이 심해져 피치 트랙을 추출하고자 하는 목적을 벗어나게 되었다.

분석 차수는 대상 유성음 구간의 피치 값의 개수가 20개 이상일 경우 매 20씩 증가시마다 1씩 증가하는 것으로 정하였다. 차수는 일반적으로 영어 유성음 구간의 길이가 평균 200-300ms라는 것을 참조하고 유성음 구간의 길이가 길어질수록 피치의 변동성이 비례하여 증가할 것이라는 가정하에 경험적으로 설정하였다.

이렇게 구한 곡선으로부터 원래 값의 개수와 같은 수의 표본을 추출하여 원래 값과의 차이를 구한 다음 피치의 변화가 제거된 평탄화된 피치의 궤적을 구한다.

그림 2는 구간내의 실제 피치값의 변동, 부족 보간에 의한 값, 오버핏된 보간에 의한 곡선, 그리고 피치 변동을 제거한 변동 곡선을 예시한 것이다. 적용된 차수에 따라 다른 곡선이 다항식 보간에 의해 구해지며 유성음 구간의 피치값의 수에 따라 적합

한 차수가 선택되고 선택한 차수에 따라 근사화된 곡선을 차감한 것이 맨 아래의 곡선으로 적용한 구간내의 피치 변동이 제거된 피치 변동 곡선이다.

연속 발화의 경우 분석 구간내에서도 피치값이 일정 범위로 변동이 심한 경우는 유효한 jitter 파라미터를 구하기 어려우므로 제외하였다.

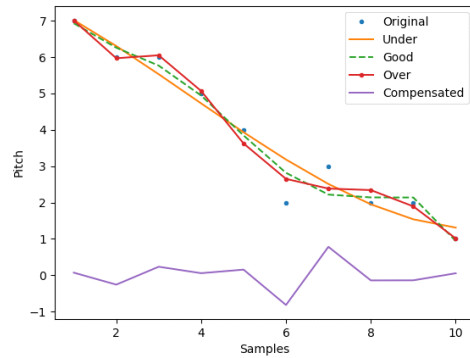


그림 2. 다항식 보간법에 의한 피치트랙 보정
Figure 2. Compensating pitch track by polynomial interpolation

그림 3은 피치 세그먼트 내에서 트랙의 변동을 4가지 유형으로 나누는 것이다. 유형 1과 2는 완만한 변동을 보여주는 경우로 변동 경향을 잘 추정할 수 있다. 그러나 유형 3과 같이 단구간에 피치값이 바뀌는 경우는 저차의 다항식으로 근사화 하기 어렵게 된다.

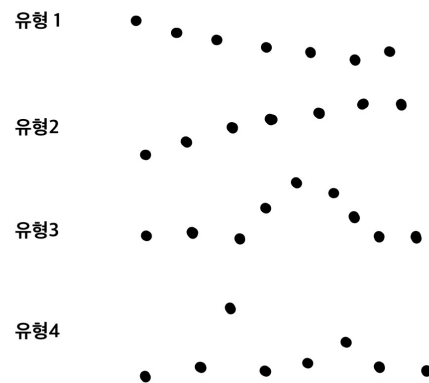


그림 3. 피치 세그먼트내의 궤적 유형
Figure 3. Types of track in pitch segment

그리고 유형 4와 같이 한 샘플씩 튀는 경우도 가끔 발생하는 데 연산 과정에서의 오차일 수도 있고 실제로 연속 발성 과정에서 짧은 구간 내에서 급격한 변동이 발생하는 경우가 있다. 유형 3과 유형 4는 특별한 경우로 취급하여 전체 jitter 계산 과정에서는 제외하는 것도 한 방법이다. 그러나 연속 음성의 특징을 나타낸다고도 볼 수 있으므로 응용분야에 따라 오히려 문장 내에서의 변동성 검출을 위하여 이 특성이 유용하게 활용될 수도

있을 것이다. 그림 4, 그림 5, 그림 6은 실제 데이터에서의 유형별 사례이다.

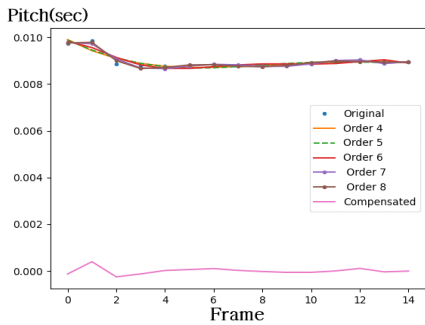


그림 4. 피치 트랙 유형 1
Figure 4. Pitch track type 1

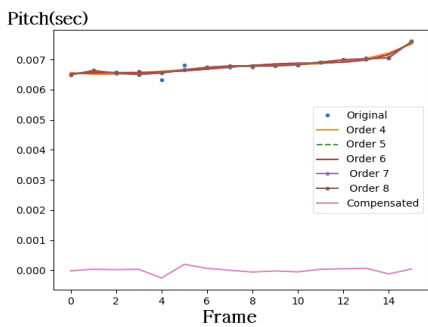


그림 5. 피치 트랙 유형 2
Figure 5. Pitch track type 2

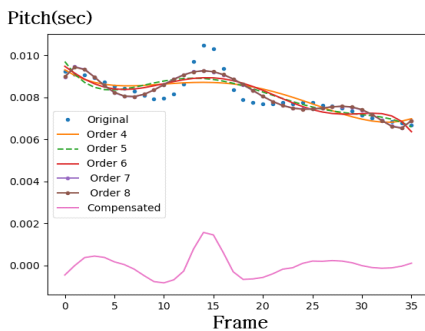


그림 6. 피치 트랙 유형 3
Figure 6. Pitch track type 3

3. 실험 및 결과

3.1. 데이터

실험에 사용한 데이터는 Kay PENTAX사의 Massachusetts Eye and Ear Infirmary(MEEL, 1994) Voice Disorders Database의 일부를 이용하였다. 이 데이터베이스에는 정상인의 음성과 각종 질환별 데이터를 포함하고 있는데 본 실험에서는 이 중 주 분석용으로 정상음성 15개와 대조분석용으로 장애음성군에서 Edema(19개), Polyp(9개)를 이용하였다. 각 데이터는 화자별로

동일 화자에 대해 3초간의 지속모음 /ah/와 12초 길이의 rainbow speech가 16 bit, 25 KHz로 표본화되어 녹음되어 있다. 본 실험에서는 rainbow speech를 분석 대상으로 이용하고 지속모음 /ah/는 화자의 기본 특성을 참조하기 위해 이용하였다.

3.2. 분석결과

실험 결과 정상음성, 장애음성에서 각각 /ah/발성과 rainbow speech에 대한 프레임별 측정치, 그리고 프레임별로 측정된 jitter값의 평균 값을 그래프로 그리고 비교하였다.

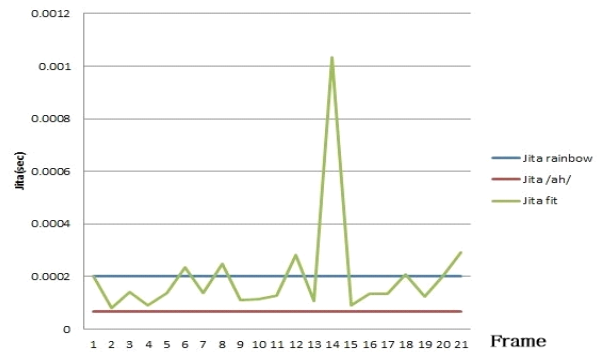


그림 7. 문장 음성에서 유성구간에서의 Jita변화
Figure 7. Jita variation in voiced segment from a sentence

표 1. 연속음성에서 유형별 정상 화자의 Jita값의 비교(p<0.05)
Table 1. Statistics of Jita in Connected Voice from normal speaker(p<0.05)

대상	Jita (msec)	
	평균	표준편차
/ah/	0.070	0.010
Rainbow	0.245	0.100

그림 7은 한 화자의 정상음성의 rainbow speech로부터 측정된 Jita값의 변화를 그래프로 그린 것이다. 표 1과 같이 일반적으로 레인보우 음문장음성 내에서 측정된 값에 비해 지속모음 /ah/에서 측정된 Jita값보다 높은 값이 측정되었다. 이것은 일반적으로 지속모음에 비해 문장음성에서 피치 변동이 증가한다는 현상과 일치한다. 제안한 방법에 의해 측정된 Jita값의 통계값은 지속모음 /ah/의 발성에 의한 통계값과 비교했을 때 ($p<0.05$)로 유의미한 변화 특성을 보였다. 동일 발성 내에서의 jitter값의 변화도 발성의 유형에 따라서 급격히 변화하는 부분이 존재하는 것을 볼 수 있었다. 급격한 변동이 존재하는 부분은 앞서 정의한 유형 3과 유형 4에 해당하는 구간이며 화자에 따라 변동이 발생하는 위치도 다름이 관찰되었다.

그림 8은 유성음 구간내에서 추출된 결과 피치트랙과 차수별 추정 잔차의 변화를 나타낸다. 그림 8(b)에서 차수가 높아질수록 잔차 오차는 줄어들지만 여기서는 구간 내에서의 피치 값의 개수에 따라 짧은 구간이므로 4차로 정하였다. 그림 8(c)는 분석 구간 내에서 차수에 따른 Jita값의 변화를 보여준다. 맨 앞의 원 신호의 값, 이후 차수가 4에서 8까지 증가할 때 다항식 보간에

의해 추정된 신호의 Jita값, 그리고 맨 마지막은 보정된 트랙에 대해 측정된 Jita값을 보여준다. 이 사례에서는 보정된 트랙에 대해 원래 신호보다 약간 줄어든 Jita값을 나타내고 있다. 이러한 경향은 동일한 문장발성 내에서도 구간에 따라 어떤 구간은 값이 증가하는가 하면 어떤 구간에서는 값이 감소하는 다른 양상을 나타내었다.

결과에 대하여 참고문헌에서 사용한 기존의 방법들은 문장 음성의 모음 지속 구간에 대해 특별한 전처리 없이 기존의 파라미터를 그대로 측정하였으므로 운율의 효과를 제거하였다고 보기 어려워 결과를 비교할 대상으로 보기 어렵다. 그래서 동일한 문장음성 내에서 전처리 없이 구한 파라미터 값과 제안한 방법에 의해 처리한 후 구한 값을 비교하였다. 여기서는 Jita값과 참고문헌에서 사용한 비교적 넓은 구간의 변동성을 포함하고 있는 파라미터인 PPQ값을 이용하여 다항식 보간 방법 적용 전과 후의 값의 차이를 비교하였다.

그림 9에서는 다항식 보간 적용 전후의 Jita값을 그림 10에서는 PPQ값을 비교하였다. 이 화자의 경우 전반적으로 적용 후 수치 값이 증가하였으나 부분별 값의 분포에 있어서는 변동에 차이를 보이고 있다. 이것은 부분적으로 운율변동의 영향이 제거 혹은 강화된 때문으로 보인다. 그 구체적인 원인에 대해서는 향후 추가적인 연구가 필요해 보인다. 분명한 것은 다항식 근사에 의한 트랙의 평탄화 과정에 의해 변동성분이 강조되어 고주파 성분이 강조되었다는 것을 확인할 수 있다. 그림9에서의 경우 동일 화자에 대해 지속모음의 경우 jita 값은 0.013ms이었고 레인보우 문장 음성의 경우 jita의 평탄화 전의 평균 값은 0.1ms로 10배 정도 증가한 값이 측정되었고 평탄화된 후 평균 값은 0.18ms로 1.8배 정도 증가하여 평탄화에 의해 불규칙성이 강조되는 효과가 관찰되었다. 그러나 문장음성의 종류에 따라 구간별로 오히려 값이 감소하는 경우도 있어 향후 문장내 음성의 구성과 종류에 따른 변화를 반영한 추가 실험이 필요하다.

그림 11과 그림 12에서는 두 명의 개별 화자에 대해 3개의 파라미터 변동을 비교한 그래프이다. 동일 화자의 경우 Jita, Jitt, RAP등 서로 다른 파라미터들의 변화 양상이 다르다는 것을 알 수 있다. Jita에 비해 더 넓은 구간의 피치값을 참조하는 RAP의 경우 변동을 보이는 구간이 다르다. 그러므로 다항식 보간에 의한 측정을 하더라도 여전히 서로 다른 파라미터에 의한 보완적 분석이 필요하며 이들 그래프를 통해 화자간 성대 진동 패턴의 차이를 확인할 수도 있다.

표 2는 제안한 방법에 의해 분석한 정상 음성과 장애음성의 Jita 값의 통계 분포를 비교한 것이다.

표 2. 연속음성에서 Jita값의 변동 특성
Table 2. Statistics of Jita in Connected Voice

파라미터	정상		장애음성	
	Jita(msec)	평균	0.245	평균
	표준편차	0.1	표준편차	3.27
	Max	0.46	Max	18.04
	Min	0.12	Min	0.14

비교에 사용한 장애음성은 MEEI DB의 Edema(19개)와 Polyp(9개)의 rainbow음성을 사용하였다. 이들 질환군은 성대 부위의 질환으로 운율 정보에 대해 최소한의 영향을 준다고 판단하여 비교 대상으로 삼았다. 평균 피치값은 장애음성의 경우 월등히 큰 값을 나타내었다. 사용한 장애음성군은 지속모음의 경우 통계적으로 정상음성과 충분히 유의미한 변별도를 갖기 어렵지만 향후 트랙에 의한 식별에 의해 장애음성 식별을 기대할 수 있다.

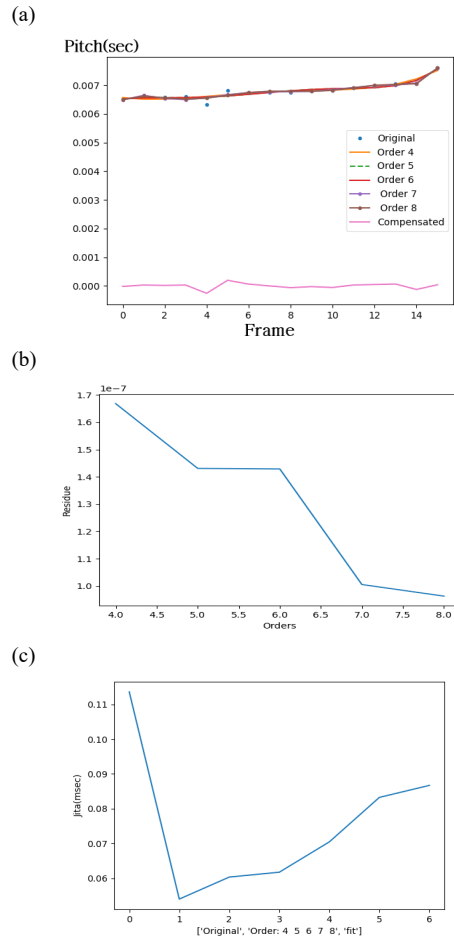


그림 8. (a)결과 피치트랙 (b) 잔차 (c) 차수에 따른 Jita 변화
Figure 8. (a) Pitch track (b) Residue (c) Jita vs order

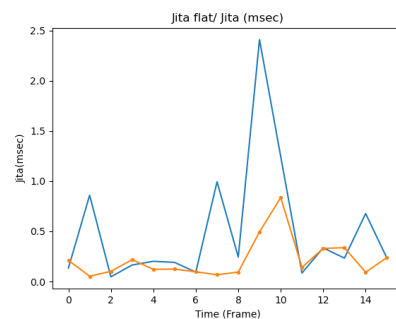


그림 9. 다항식 보간법 전(점)후의 Jita값의 변화
Figure 9. Jita before(dot) and after flattening

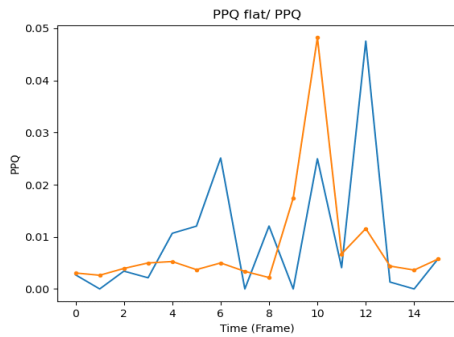


그림 10. 다항식 보간법 전(점)후의 PPQ값의 변화
Figure 10. PPQ before(dot) and after flattening

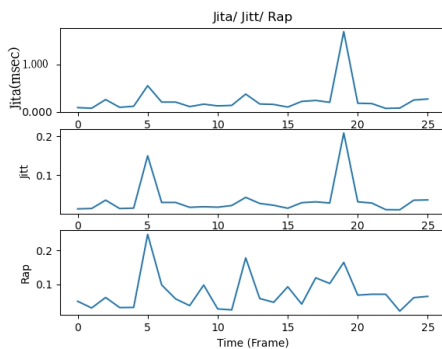


그림 11. 레인보우 음성 파라미터 트랙 A
Figure 11. Rainbow speech parameter track A

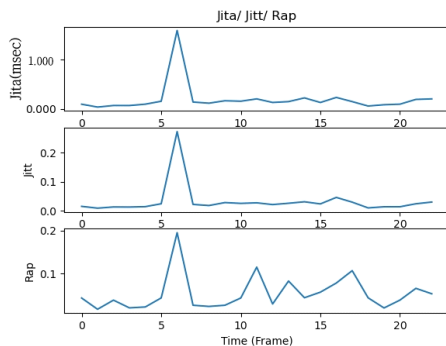


그림 12. 레인보우 음성 파라미터 트랙 B
Figure 12. Rainbow speech parameter track B

4. 결론

본 논문에서는 연속발성 문장음성으로부터 피치의 변동성을 측정하기 위하여 다항식 보간에 의한 피치계적 평탄화 방법을 제안하고 이 방법에 의해 운율 정보가 제거된 피치 계적으로부터 연속음성의 피치 변동 파라미터를 구하는 과정을 보였다.

다항식 보간을 통해 구한 jitter파라미터는 정상음성군의 지속 모음 /ah/의 경우와는 값의 범위에서 차이가 있었지만 동일한 수준의 통계값 변화를 보여줌을 확인하였고 음성의 음질 측정 및 정상음성과 장애음성간 식별 파라미터로서의 활용 가능성

을 보여주었다. 또한 레인보우 음성을 이용한 실험에서 원 음성 과 비교하여 증가된 변동성 및 구간별 변동성의 변화를 관찰할 수 있었다.

향후의 연구 과제로는 궤적의 평탄화에 따른 구간별 특성 변화 측정 및 트랙에 의한 장애음성 식별에의 적용 등이 있다. 또한 제시한 방법을 개선하여 연속발성 문장 음성을 통한 음성의 분석 및 식별에 관한 연구에 적용하고자 한다.

References

- Fourcin, A. (2009). Aspects of voice irregularity measurement in connected speech. *Folia Phoniatrica et Logopaedica*, 61(3), 126-136.
- Gerratt, B. R., Kreiman, J., & Garellek, M. (2016). Comparing measures of voice quality from sustained phonation and continuous speech. *Journal of Speech, Language, and Hearing Research*, 59(5), 994-1001.
- Kisenwether, J. S., & Prosek, R. A. (2017). Long-term and short-term period and amplitude perturbation measurements: Are they all needed? *Annals of Otolaryngology and Rhinology*, 4(4), 1172.
- MacCallum, J.K., Zhang, Y., Jiang, J.J. (2011). Vowel selection and its effects on perturbation and nonlinear dynamic measures. *Folia Phoniatrica et Logopaedica*, 63(2), 88-97.
- Manfredi, C., Giordano, A., Schoentgen, J., Fraj, S., Bocchi, L., & Bejonckere, P. H. (2011). Perturbation measurements in highly irregular voice signals: Performances/validity of analysis software tools. *Biomedical Signal Processing and Control*, 7, 409-416.
- Massachusetts Eye and Ear Infirmary. (1994). *Voice disorders database (Version 1.03 CD-ROM)*. Lincoln Park, NJ: Kay Elemetrics.
- Schoentgen, J. (1989). Jitter in sustained vowels and isolated sentences produced by dysphonic speakers. *Speech Communication*, 8(1), 61-79.
- Schoentgen, J., & Guchteneere, R. (1995). Time series analysis of Jitter. *Journal of Phonetics*, 23(1), 189-201.
- Teixeira, J.P., & Golçalves, A. (2014). Accuracy of jitter and shimmer measurements. *Procedia Technology*, 16, 1190-1199.
- Vasilakis, M., & Stylianou, Y. (2009). Voice pathology detection based on short-term jitter estimations in running speech. *Folia Phoniatrica et Logopaedica*, 61(3), 153-170.

- 조철우 (Cheolwoo Jo) 교신저자
창원대학교 전기전자제어공학부 교수
경남 창원시 의창구 창원대학교 1
Tel: 055-213-3662
Email: cwjo@changwon.ac.kr
관심분야: 음성신호처리, 장애음성 진단 및 처리

음성신호의 Jitter 성분의 장시간 변화에 관한 통계적 분석*

조 철 우

창원대학교 전기전자제어공학부

국문초록

본 연구에서는 연속음성에서의 장시간 jitter 성분 측정 방법에 대해 고찰하였다. 기존의 jitter 측정방법으로는 지속 발성한 모음을 대상으로 변동성을 측정하는 방법을 주로 사용하여왔다. 문장음성 등 연속음성의 경우는 문장에 따른 운율정보의 영향으로 기존의 측정법으로는 왜곡이 발생하게 된다. 이에 연속 발성에 대해 운율정보의 피치 변동을 상쇄시키는 방법을 제안하고자 한다. 피치 변동을 제거하는 방법으로는 분석구간내에서의 피치 변동을 다항식 보간법에 의해 변동 경향을 대표하는 곡선을 구하고 그 곡선을 기준으로 변이를 제거하였다. 이후 변이가 제거된 피치의 궤적으로부터 jitter를 측정하는 방법을 적용하여 피치 주파수의 변동성을 측정하고 기존의 지속모음에 의한 측정 방법과 비교하였다. 제안한 방법의 효용성 측정을 위해 Kay Pentax MEEI DB의 음성 표본을 사용하였다. 통계분석 결과 제안된 방법에 의해 연속음성으로부터 측정된 jitter 값은 동일 화자의 지속모음으로부터 측정된 파라미터 값과 유사한 변동성을 보여 주었다.

핵심어: Jitter, 연속음성, 변동성 측정, 포물선 보간법, 지속모음

* 이 논문은 2019-2020년도 창원대학교 자율연구과제 연구비 지원으로 수행된 연구결과임.