

음성과형의 평균진폭 변동율에 의한 주기별 피치검출

강동규, 정유현, 최준혁
한국전자통신연구소 음성응용연구실

Period-to-Period Pitch Estimation Using Average Magnitude Fluctuation Rate

KANG DONG GYU, JEONG YOU HYEON, CHOI JOON HYUK
ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE
SPEECH APPLICATION SECTION

요 약

선형예측 기법으로 피치동기식 분석을 하기 위해 우선적으로 필요한 정보로서 유성음 신호의 영교차 구간별 평균 진폭변동을 추출에 의한 주기별 피치를 검출할 수 있는 기법을 제안하였다.

유성음의 제 1포먼트 성분에 대한 각 영교차 구간에서의 평균진폭값은 성대 폐쇄시점에서 주기별 최대치를 나타내며, 평균진폭변동율은 "+" 영역의 평균진폭값과 선행하는 "-" 영역 값의 차로 표시한다. 이 평균 진폭 변동율은 성대파형의 영향이 반영되어 주기성이 더욱 강조되므로 분석구간에 대한 구간별 평균피치와 변화의 정도를 이용하여 주기별 피치정보를 추출할 수 있다.

검출결과는 구간별 평균 피치와 비교하였으며, 좋은 결과가 나타나는 것을 확인할 수 있었다.

1. 서론

널리 이용되고 있는 프레임 동기식 음성 분석방법은 분석구간이 짧을 경우에 음성신호의 주기성에 의해 발생하는 smearing 현상과 분석구간이 긴 경우에는 기본 주파수의 하모닉스에 의한 영향 때문에 정확한 음성의 특징 변수를 추출하기 어려운 단점을 가지고 있다. 이러한 영향을 줄일 수 있는 방법은 피치동기식 분석으로서 먼저 주기별 피치검출이 이루어져야 한다.¹⁾

정밀한 주기별 피치검출에 관한 연구는 성대 폐쇄시점인 GCI(Glottal Closure Instant)를 결정하는 문제로 볼 수 있으며, 이 GCI가 결정되면 성대폐쇄 구간의 음성신호를 zero-input 응답으로 간주할 수 있다. 따라서 이 구간에 대해 분석을 수행하면 이상적인 성도특성을 얻을 수 있고, 다시 역 필터링에 의해 음원을 분리하여 소량의 파라미터로 음성의 특징을 정밀하게 표현할 수 있다. 그러므로 정확한 GCI 검출은 음성인식, 합성 및 부호화 시스템에서 많은 성능 개선을 기대할 수 있는 동시에 정밀한 주기변화에 의한 화자정보, 음운정보 등을 추출할 수 있는 장점이 있어서 이에 관한 많은 연구가 진행되어 왔다. [1-8]

지금까지 발표된 GCI 검출 방법들로서는 LP분석의 잔차신호에 의한 방법, epoch filter에 의한 방법, auto-covariance log determinant에 의한 방법 등이 있다. 이 방법들은 대부분 성대가 완전히 닫히지 않거나 닫힌 구간이 너무 짧아 분석에 충분한 데이터를 갖지 못할 경우 많은 검출 오차를 보이고 있어서 아직까지 어려운 문제로 남아 있다. [1-8]

GCI 결정에 관한 알고리즘들은 음성신호에서 잔차신호나 epoch 신호를 추출하고 이 신호로부터 직접 GCI를 결정하기 때문에 GCI를 명확히 구별할 수 없는 구간에서는 많은 검출 오차를 나타내고 있다. 본 논문에서는 GCI에서 발생하는 가장 큰 에너지 변화를 나타낼 수 있는 평균진폭 변동율을 추출하고 분석구간별 평균 피치를 참조로 근사적인 GCI를 결정할 수 있는 기법을 제안하였다.

2. 유성음의 주기별 평균 진폭 변화

음성신호는 음원에 따라 무성음, 유성음, 혼합음으로 구분할 수 있으며 유성음의 경우 성대의 진동을 여기원으로 성도가 공명되어 발생되고 성도의 물리적인 저항성분에 의해 피치주기 내에서 GCI를 시점으로 지수함수적인 감쇄를 나타내므로 GCI에서 최대의 에너지 변화를 나타낸다. [1-8]

유성음 신호에서 제 1포먼트의 에너지는 다른 포먼트에 비해 약 10dB 이상 높으므로 시간영역에서 볼 때 가장 큰 진폭변화를 나타내어 피치주기내에서의 영교차율은 제 1포먼트의 주파수를 근사적으로 나타낸다. 그림 2-1에서 제 1포먼트에 대한 근사 분석과 같이 감쇄진동에 의해 GCI부근에서 최대치를 나타내는 피치주기별 진폭변화를 관찰할 수 있다. [1]

주기별 진폭변화 특성을 이용한 피치 검출 기법들 중에서 parallel processing 방법은 900Hz 미만의 성분에 대한 peak와 valley의 최대 및 최소치에 의해 주기성을 강조하여 피치를 검출하고 있다. [1]

이 방법은 진폭의 크기를 이용하여 주기성을 강조하므로써 그림 2-1에서와 같이 성대파형의 영향이 작은 진폭과 낮은 주파수로 나타날 경우 이를 충분히 반영하지 못

음성파형의 평균진폭 변동에 의한 주기별 피치검출

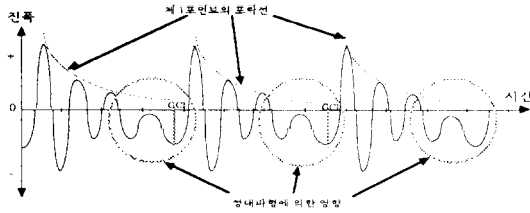


그림 2-1. 음성음의 제 1포먼트에 대한 근사 분석

할 뿐만 아니라 음성파형의 피크 정보를 이용하므로서 정밀한 주기별 피치검출에 적용하기는 어렵지만 피치의 주기결정에서는 많이 이용되고 있다.

한편 보다 정밀한 위치정보와 고속처리가 가능한 방법으로서는 면적 비교법이 제안되었으며, 이 방법은 음성음에서 제 1포먼트의 영교차구간이 가장 길고 진폭이 큰 특징을 나타낼 수 있는 변수가 영교차 구간에서의 파형 면적임에 착안한 것이다. 영교차점 위치 k_c 에 대한 면적함수는 식 (2-1)과 같이 표시할 수 있다 [1].

$$A(k_c) = \sum_{j=1}^k S(j) \quad (2-1)$$

식 (2-1)에서 "+" 영역의 면적만을 고려하면 그림 2-2와 같이 감쇄진동에 의한 현상으로 한 피치 주기내의 최대 피크값을 나타내므로 면적값들 중에서 면적의 크기와 주기를 비교하여 피치를 결정할 수 있다.

이 방법은 그림 2-2에서와 같이 "+" 영역의 면적만을 고려하고 있으나 parallel processing 방법에서와 같이 "+" 면적과 선행하는 "-" 면적의 차를 구하면 그림 2-3과 같이 제 1포먼트 성분에 대한 근사적인 영교차구간에서의 평균 진폭변화로 나타낼 수 있다.

음성신호 $S(n)$ 에서 진폭의 값이 "+"에서 "-"로 변화하는 영교차점을 zc_i 라 하고, 이 zc_i 에 선행하여 "+"에서 "-"로 변화하는 영교차점의 위치를 zc_{i-1} 라 하면, "+"영역의 면적함수 $A^+[zc_i]$ 와 "-"영역의 면적함수 $A^-[zc_i]$ 는 각각 다음의 식 (2-2), (2-3)으로 나타낼 수 있다.

$$A^+[zc_i] = \sum_{n=zc_{i-1}}^{zc_i} S(n) \quad (2-2)$$

$$A^-[zc_i] = \sum_{n=zc_i}^{zc_{i+1}} S(n) \quad (2-3)$$

여기서 $i=0, 1, 2, 3, \dots, l$ 은 분석구간내에서 "+"와 "-"영역별 영교차점의 일련번호를 나타내며 만일 zc_i 에 선행하는 zc_{i-1} 가 없을 경우에는 zc_{i-1} 를 삭제한다. 영교차점 위치 zc_i 에서의 "+" 면적과 선행하는 "-" 면적의 차를 나타내는 평균 진폭변동함수 $MF[zc_i]$ 는 식 (2-4)와 같이 표현할 수 있다.

$$\begin{aligned} MF[zc_i] &= A^+[zc_i] - A^-[zc_i] \\ &= \sum_{j=zc_{i-1}}^{zc_i} S(j) - \sum_{k=zc_i}^{zc_{i+1}} S(k) \end{aligned} \quad (2-4)$$

그림 2-4는 실제의 음성신호에 대한 각 영교차 구간

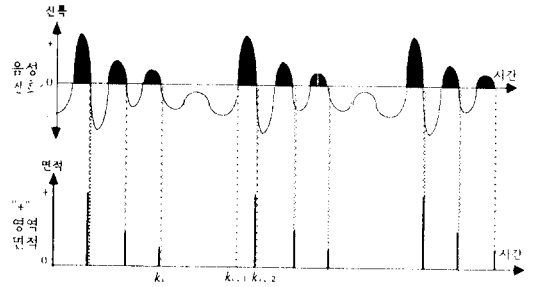


그림 2-2 면적 비교법에서 "+"영역의 면적값의 배열

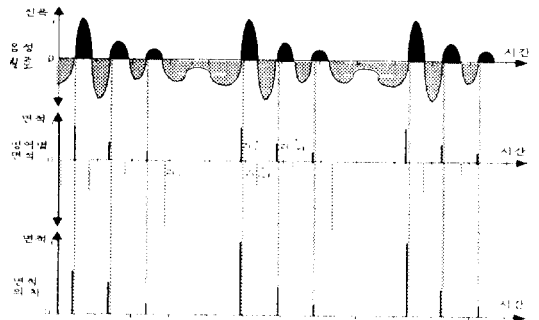


그림 2-3 면적의 차에 의한 평균진폭변화

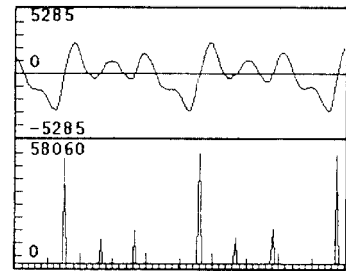


그림 2-4 음성신호에 대한 각 영교차구간에서의 평균 진폭 변동용 (상) 음성신호의 신호, (하) 영교차구간별 평균진폭변화 $MF[zc_i]$

에서의 평균 진폭변화를 나타낸 것으로 피치 주기내에서 뚜렷이 구별되는 평균 진폭변화를 나타내고 있음을 알 수 있다.

3. 주기별 피치검출

음성음은 그림 2-1에서와 같이 상대파형에 의한 "+" 영역의 큰 면적과 이후에 나타나는 첫번째 "-" 영역의 면적이 피치주기 내에서 가장 크거나, 상대파형의 영향이 적은 경우에는 첫번째 "+" 면적이 상대적으로 큰 특징을 가지므로 각 영교차 구간에서의 평균 진폭변동함수는 주기내의 GCI 부근에서 뚜렷히 구별되는 가장 큰값을 나타낸다.

각 영교차 구간의 평균진폭 변동함수 $MFI(z_c)$ 로부터 직접 GCI정보를 산출할 수도 있지만 피치의 변화를 경우에는 주기성이 적고, 또 음성의 에너지 변화가 큰 경우에는 주기내 평균 진폭변동율의 최대치가 다른 주기내의 최대치가 아닌 값들 보다 작을 수도 있어 검출에 어려움이 있다. 그러나 분석구간내의 평균피치를 알고 있다면 용이하게 피치주기내의 최대 진폭변화율을 결정하여 피치의 주기 및 위치정보를 추출할 수 있다.

먼저 분석 프레임별 평균 피치를 구하기 위해 이미 발표된 가변길이 dual impulse의 스펙트럼 하모닉스 매칭에 의한 방법을 사용하였다. 이 방법은 극성이 다른 dual impulse의 스펙트럼과 음성음의 스펙트럼이 곱해질때 나타나는 에너지 변화를 이용하고 있다. 간격 N 인 dual impulse $iaw(n)$ 과 음성신호 $s(n)$ 을 convolution하여 얻은 신호 $bs(n)$ 을 식 (3-1)과 같이 표시하고,

$$bs(n) = iaw(n) * s(n) \quad \dots (3-1)$$

분석구간이 m 이고 함수 MAX 과 MIN 을 구간 N 에서 최대치를 구하는 함수라 할때 에너지의 변화를 impulse 간격 N 으로 나타낸 에너지 변화함수 $EF(N)$ 은 식 (3-2)와 같이 표시된다.[3]

$$EF(N) = \frac{1}{m-N} \sum_{i=0}^{m-N} \left[MAX_V\{bs(n-i)\} - MIN_V\{bs(n-i)\} \right] \quad m > N \quad \dots (3-2)$$

식 (3-2)의 에너지 변화함수에서 첫번째 골과 이것의 대역폭에 의해 평균 피치와 분석 구간내에서의 피치 변화도를 추출할 수 있고 이 정보를 참조하면 평균 진폭변동함수로부터 결정논리에 의해 주기별 피치정보를 추출할 수 있다.

$MFI(z_c)$ 로부터 주기별 피치검출을 위한 결정논리는 피치의 두가지 기본 특성을 적용하여 수행될 수 있다. 첫번째로 $MFI(z_c)$ 의 값은 상대파형의 여가순간에서 큰 값을 나타내고 시간이 경과함에 따라 점점 작아지는 점이며, 두번째로서는 이러한 현상이 준주기적으로 반복된다 는 점이다.

첫번째 성질을 적용한 결정논리는 먼저 평균피치검출기로부터 산출된 분석 구간내의 최소피치에 의해 탐색구간 min_pth 를 결정한다. $MFI(z_c)$ 에 대하여 샘플단위로 이동하면서 탐색구간 min_pth 내에서 peak picking에 의해 최대치를 검출하고 이 보다 작은 값들을 0으로 대체하면 $MFI(z_c)$ 에는 대체로 피치 주기내의 최대 평균진폭 변동 값들만 남게 된다. 이때, z_c 는 영교차점의 의미보다는 $MFI(z_c)$ 가 영 이상의 값을 가지는 영교차점을 나타낸다.

준주기적인 성질을 이용하는 두번째 결정논리는 평균 피치검출기에서 산출한 최소와 최대 피치값에 의해 주기성을 판별하는 것이다. $MFI(z_c)$ 에서 (z_c^1, \dots, z_c^N)는 영보다 큰 평균진폭 변동값이 존재하는 인접 영교차 점과의 간격이므로 이 값이 최소피치 min_pth 와 최대피치 max_pth 의 범위에 속하는지 비교한다. 만일 이 범위를

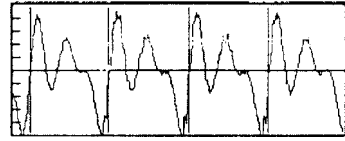


그림 3-1 검출된 주기별 피치의 음성신호

벗어나는 경우에는 피치의 주기성이 없는 것으로 판정하고 평균진폭 변동율 $MFI(z_c)$ 를 0으로 대체하며, 위의 과정을 샘플단위로 반복 수행한다.

두번째 결정논리에 적용된 피치의 변화도는 군사적인 값이므로 분석구간 내에 존재할 수 있는 최적의 GCI수 max_gci 가 추출될 때까지 비교범위를 적용하기 위해 먼저 분석구간 m 내에 존재할 수 있는 max_gci 를 식 (3-3)과 같이 결정한다.

$$max_gci = \frac{m}{min_pth} \quad \dots (3-3)$$

min_pth 와 max_pth 를 매회마다 각각 증가 및 감소시켜 비교범위를 좁히면서 두번째 결정논리를 반복 수행하고 GCI 개수를 나타내는 i 이 max_gci 보다 작거나 같을 경우 최적의 주기별 피치들을 추출한 것으로 판정하고 결정논리 과정의 수행을 종료한다.

이상의 두가지 결정논리를 통과한 $MFI(z_c)$ 에서 z_c^i 는 군사적인 GCI정보를 나타낸다. 처리 시간의 단축을 위해 두 단계의 결정논리에서 행해지는 비교연산을 비교구간내의 z_c^i 에서만 수행할 수 있다.

그림 3-1은 실제의 음성신호에 대하여 평균 진폭변동율에 의한 주기별 피치검출 결과(수직선)를 음성신호와 비교한 것으로 검출결과를 군사적인 GCI를 나타내고 있음을 볼 수 있다.

4. 실험 및 결과

$MFI(z_c)$ 를 산출하고 2단계의 결정논리 수행에 대한 컴퓨터 simulation을 하기 위해 30대 남녀가 아래와 같은 문장을 발성 한 다음 10KHz로 표본화하면서 16-bit A/D 변환을 수행하여 저장하였다.

발성1) 남성화자 " 그리고 우리를 간절히 부르고 있다."
발성2) 여성화자 " 인간은 어느 한쪽 되만 가지고는"

먼저 각 음성시료에 분석 길이를 평균피치 검출기에서 구한 최대피치의 2배 이상(2~3.6msec)이 되도록하고 100샘플씩 중첩시키면서 900Hz로 저역통과를 한후 분석을 수행하였다. 각 영교차 구간에서 평균진폭 변동율을 추출하여 분석 프레임내의 평균 피치를 참조로 평균 진폭 변화함수로부터 결정논리에 의해 주기별 피치정보를 검출하였다. 그림 4-1, 4-2에서 실선으로 표시된 평균 피치 값에 점선으로 검출결과를 비교하였으며 검출결과는 평균 피치를 중심으로 약간의 편차를 나타내고 있음을 볼 수 있다.

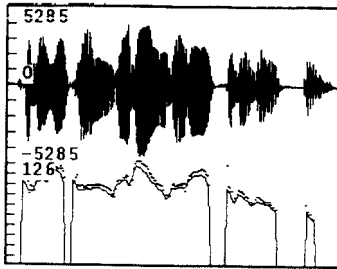


그림 4-1. 발성1) "그리고 우리를 간절히 부르고 있다."에 대한 처리 결과

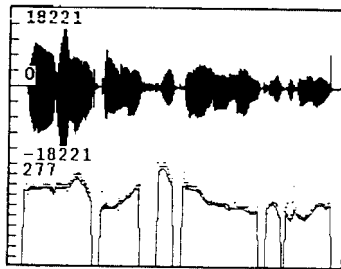


그림 4-2. 발성 2) "인간은 어느 한쪽 뇌만 가지고는"에 대한 처리 결과

5. 결론

정확한 GCI검출이 가능하면 음원의 분리가 용이할 뿐 아니라 정밀한 성도 특성을 얻을 수 있으므로 음성관련 분야에서 많은 관심을 가지고 연구가 진행되어 왔다.

지금 까지 알려진 방법들은 음성신호에서 추출한 잔차 신호나 epoch 신호에서 직접 GCI를 검출하고 있기 때문에 GCI를 명확히 구분하기 어려운 경우에는 많은 검출 오차를 보이고 있다. 본 논문에서는 GCI에서 발생하는 최대 에너지변화를 나타낼 수 있는 평균진폭 변동율을 추출하여 분석 구간별 평균피치를 참조로 근사적인 GCI를 결정할 수 있는 기법을 제안하였다.

각 영교차 구간에서의 평균 진폭변동율은 성대 파형의 영향을 보다 잘 나타내므로 GCI부근에서 뚜렷한 피크를 보여 분석구간별 평균 피치를 적용하면 쉽게 주기별 피치를 결정할 수 있다. 이 방법은 주로 덧셈과 비교연산만이 필요하므로 고속처리가 가능할 뿐 아니라 주기별 피치의 변화폭이 작은점과 이전 분석구간의 피치값들을 적용하여 평균 피치검출기에서 발생한 검출 에러를 보정을 할 수 있는 장점도 지니고 있다.

제안된 방법은 교역성분에 의한 영향을 줄이기 위해 900Hz로 저역통과 필터링을 하였기 때문에 실제의 음성신호와와는 위상차로 인한 오차를 나타내고 있어서 정밀한 GCI를 결정하기 위한 연구를 현재 진행중에 있다. 동시에 이를 바탕으로 피치동기식 분석기법에 관한 연구도 진행중에 있다.

감사의 글

본 논문은 한국통신 출연과제인 "통신처리 시스템의 고급기능 개발 과제"중 동경대와의 국제공동연구인 "피치동기식 포먼트 추출에 관한 연구"로 수행된 결과물입니다. 원활한 연구를 위해 많은 도움을 주신 윤병남 부장님과 음성응용연구실 실험들에게 감사드립니다.

참고 문헌

- [1] 배명진, 안수길, 면적 비교법에 의한 고속 PITCH 추출, 대한전자공학회지, 제 22권, 제 2호, 1985.3.
- [2] 강동규, 배명진, 안수길, "배경잡음하에서 스펙트럼 누설현상을 이용한 음성신호의 중심 피치검출," 한국음향학회 논문집 10권 1호 pp.37~46, 1991.2.
- [3] 강동규, 정유현, 가변 길이 Dual Impulse의 Spectrum Harmonics Matching에 의한 피치검출에 관한 연구, 제 7회 신호처리 종합학술대회 논문집, 제 7권, 1호, PP. 775~778, 1994.10.
- [4] Yan Ming Cheng, Douglas O'shaughnessy, Automatic and Reliable Estimation of Glottal Closure Instant and Period, IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-37, No. 12, pp. 1805~1815, December 1989.
- [5] T. V. Ananthapadmanabha, B. Yegnanarayana, Epoch Extraction from Linear Prediction Residual for Identification of Closed Glottis Interval, IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-27, No. 4, pp. 309~319, August 1979.
- [6] T. V. Ananthapadmanabha, B. Yegnanarayana, Epoch Extraction of Voiced Speech, IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-23, No. 6, pp. 562~570, December 1975.
- [7] Hans Werner Strube, Determination of the Instant of Glottal Closure from the Speech Wave, J. Acoust. Soc. Am., Vol. 56, No. 5, pp. 1625~1629, December 1974.
- [8] S. Parthasarathy, Donald W. Tufts, Excitation-Synchronous Modeling of Voiced Speech, IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-35, NO. 9, pp. 1241~1249, September 1987.
- [9] L.R. Rabiner/R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Inc. Englewood Cliffs, New Jersey, pp.98~105, pp.135~141, 1978.
- [10] 김재인, 이종락, 성문 폐쇄 시점 검출에 관한 연구, 제 10회 음성통신 및 신호처리 워크샵 논문집, 제 SCAS-10권, 1호, PP.170~172, 1994.8.