

SHS 방법과 P&G 방법의 조합을 통한 자동 음고 추정

김동수, 방희석, 백성준, 성평모
서울대학교 전기공학부

Automatic Fundamental Frequency Detection by Mixed SHS and P&G Method

Dong-Soo Kim, Hee-Suk Pang, SeongJoon Baek, Koeng-Mo Sung
School of Electrical Engineering, Seoul National University

e-mail : cesibon@acoustics.snu.ac.kr

요약

본 논문에서는 악기음의 음고 추정을 위해 많이 사용되어지는 것으로 알려진 SHS 방법과 P&G 방법의 조합 및 개선을 통하여 좀더 신뢰도가 높은 음고 추정 방법을 제안한다. P&G 방법의 경우 잡음에 의한 주파수 조합의 오염에 민감한 단점이 있는데 주파수 조합을 선택함에 있어 SHS 방법을 사용하여 잡음에 의한 주파수 조합의 오염을 비교적 무시할 만한 수준으로 낮출 수 있었다. 또한 P&G 방법에 있어 배음 행렬 개념을 도입하여 음고 추정시 가장 빈번히 발생하는 오류인 옥타브 에러(octave error)를 최소화하였다. 또한 본 알고리즘을 여러 악기음에 적용한 결과, 기존 방법에 비해 더 나은 성능을 보였다.

1. 서론

악기음이나 음성은 기본주파수의 배수인 배음들(harmonics)로 이루어져 있다. 이러한 배음들 중 기본주파수음과 음고를 찾는 연구는 오랫동안 진행되어 왔다. 제안된 방법들은 크게 두 가지로 나눌 수 있는데, 첫째는 스펙트럼상의 피크들을 찾은 후 이들로부터 기본주파수 음으로서의 확률이 가장 높은 것을 찾는 방법이고, [1,2] 둘째는 스펙트럼 자체에 적절한 후처리를 하고 그 결과로 주파수 음을 찾는 방법이다. [3,4] 이들 방법의 대표적인 예로는 P&G(Piszczalski and Galler's component frequency ratio) 방법 [1]과 SHS(Sub-Harmonic Summation) 방법 [3]이 있다.

P&G 방법은 FFT에서 피크들의 주파수와 그 크기를 뽑아낸 후, 각각 두 개씩의 주파수를 비교하는 방법을 통해 기본 주파수를 추정하는 방법이다. 피크로 결정된 주파수의 개수가 n 개인 경우 주파수 정수비의 최고값인 12를 행의 값으로 하고 주파수 개수 n 을 열로 하는

$12 \times n$ 행렬을 만들게 된다. 이 행렬에 각각의 주파수 쌍의 비교를 통해 가중치를 더하게 된다. 이와 같은 방법으로 선택되어진 모든 주파수의 쌍들에 대해 값들을 대입한 후에 행렬에서 가장 큰 값을 갖는 곳에 해당하는 값을 기본주파수로 정하게 된다.

P&G 방법의 경우 여러 가지 문제점을 갖고 있는데 그 중에 대표적인 경우가 미리 정해진 정수비가 필요 이상으로 너무 조밀하여 결과치를 전체적으로 산만하게 만드는 요인이 되며 또한 2:1과 같은 경우 두 주파수의 비가 2:1인지, 아니면 4:2인지 명확히 구분할 수 없다는 문제점을 안고 있다. 또한 4:3과 같은 일반적인 정수비의 경우 각각의 주파수에 해당하는 열의 4행과, 3행에 값이 더해지게 되는데, 이러한 방법으로 더하는 것이 결국에는 가중치 값의 분산을 초래하여 정확한 음고를 추정할 수 없게 한다. 또한 음고와 상관없는 영역의 잡음 성분이 여럿 있는 경우 이의 영향이 무시할 수 없는 정도로 커지게 되어 결국 잡음 성분이 기본 주파수로 잘못 인식 되는 경우가 있다.

SHS 방법은 음고가 존재하는 신호의 경우 FFT의 피크 값들이 주파수의 linear scale 상에서 기본 주파수의 배음 형태 즉, 일정한 간격을 두고 존재하는 성질을 이용하여 음고 추정을 한다. linear scale 상에서 일정한 간격을 두고 존재하는 FFT의 피크 값들을 순차적으로 압축한 값들을 모두 더하게 되면 기본주파수에 해당하는 값에서 배음들이 순차적으로 더해지게 되어 결과적으로 가장 큰 값을 얻을 수 있게 된다. 이러한 SHS 방법의 경우 기본 주파수 음이 없는 경우에 있어서도 기본 주파수를 쉽게 찾을 수 있는 장점이 있다. 하지만 알고리즘 중의 spectral window 함수와 가중치 값을 최적화 시킴에도 불구하고 음고 추정에 있어 가장 많은 에러의 요인인 옥타브 에러가 발생하는 것을 막을 수 없는 단점이 있다.

II. 제안된 방법

제안된 방법의 흐름도를 그림 1에서 보이고 있다.

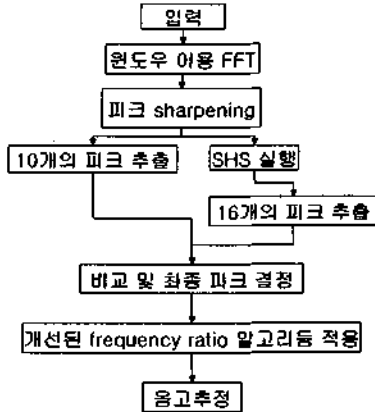


그림 1. 제안된 방법의 흐름도

우선 신호는 프레임 별로 해닝 윈도우가 씌워진 후 FFT가 수행된다. 그 후에 스펙트럼에 대해 피크 샤프닝을 하는데 이는 후에 이루어질 SHS에서 같은 메인 로브의 값들이 중첩되는 것을 방지함과 동시에 잡음에 의한 영향을 배제하기 위해서이다. 피크 샤프닝은 다음과 같은 절차로 이루어진다. FFT의 k 번째 값을 p_k 라 할 때,

$$p_k > p_{k-1} \text{ and } p_k > p_{k+1} \text{ and}$$

$$p_k^2 + p_{k-1}^2 + p_{k+1}^2 \geq 10(p_{k-2}^2 + p_{k+2}^2) \text{ then}$$

$$p_{k-2} = p_{k-1} = p_{k+1} = p_{k+2} = 0$$

그림 2-(a)와 같은 FFT 결과에 대해 그림 2-(b)에서 피크 샤프닝을 수행한 결과를 보이고 있다. 이제 이를 다음과 같은 절차로 SHS를 수행한다.

$$Y(n) = \sum_{i=1}^8 X(i \cdot n)$$

where

$$X(k) = 0, \quad k \leq 3$$

$$X(k) = \max[X(k-1), X(k), X(k+1)], \quad k \geq 4$$

단 $X(k)$ 는 신호의 FFT 크기 값을 의미하며 이를 그림 2-(d)에 보이고 있다. 이제 그림 2-(c) 및 2-(e)에서 FFT에서 10개, SHS 된 결과에서 16개의 피크를 뽑아낸 그림을 보이고 있다. SHS 결과에서 더 많은 피크를 뽑는 것은 각 배음의 1/2, 1/3 등의 위치에서 피크가 생기므로 이를 고려해서이다. 이제 이들의 결과를 비교하여 공통된 위치에 존재하는 피크들을 2-(f)에서 보인다. 원래의 FFT 결과에서만 피크를 추출하면 잡음에 의한 피크가 추출되어 전체 성능에 영향을 줄 수가 있다. 또

한, SHS 방법에서는 배음의 1/n 지점에 피크가 생겨서 성능이 나빠진다. 따라서 이 들을 모두 고려하면 잡음에 의한 피크와 배음의 1/n 위치의 피크들의 영향을 배제할 수 있어서 실제 배음들만 피크로 추출할 수 있게 된다.

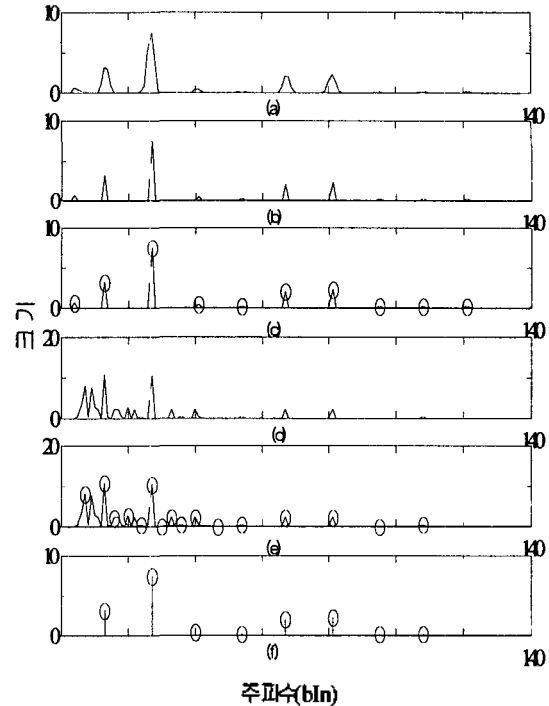


그림 2

이제 마지막으로 추출된 피크를 이용하여 음고를 추정하게 되는데, 여기에서는 기존의 P&G 방법에서 사용하는 frequency ratio 알고리즘을 개선한 것을 사용하였다. 이를 위한 수정방법은 표 1에 간략히 나타내었다.

표 1. (f_a, f_b) 의 가중치 계산법 비교

| | 기존방법 | 개선된 방법 |
|-------------|-------------------------|---|
| 주파수비가 1:k일때 | m:mk 모두에 가중치 부여 | m:mk중 $f_a/m, f_b/mk$ 에 피크가 있는 곳에만 가중치 부여 |
| 주파수비가 i:j일때 | a의 i항과 b의 j항에 각각 가중치 부여 | i와 j로부터 추정되는 기본 주파수에만 가중치 부여 |
| 옥타브 에러 검사 | 하지 않음 | 배음행렬을 도입하여 1/n 옥타브 에러 검사 |

위의 표에서 새로 도입한 배음행렬에 대하여 좀더 자세히 설명하도록 하자. 배음행렬은 초기에는 모든 성분값을 0으로 갖는 행렬이다. 두 개의 주파수 성분을 각각 비교하는 경우에 있어 두 주파수 f_a 와 f_b 의 정수비가 a:b의 정수비를 갖고, 이에 해당하는 기본주파수가 존재하는 경우 P&G 방법에서는 행렬에 가중치를 더하게 된다. 이와 동시에 배음행렬이라는 새로운 행렬을 도입하여 이곳에는 기본주파수에 해당하는 행의 a번째와 b번째의 열의 성분에 '1'값을 할당한다. 이는 해당 기본주파수의 a번째 배음과 b번째 배음이 존재함을 알게 해준다. 이러한 '1'값의 할당은 결과적으로 P&G 방법의 행렬 즉, 가중치 값을 더하는 행렬에서의 가장 큰 값을 갖는 곳에서의 주파수를 기본주파수로 결정하기에 앞서 기본주파수로 결정하는 것이 적합한지를 검증할 수 있게 해준다. 가중치값의 합이 가장 큰 곳에 해당하는 주파수가 기본주파수인 경우 이에 해당하는 배음행렬의 행은 연속적인 1 즉, [1 1 1 1 1 ...]과 같은 형태를 띄게 된다. 반면 기본 주파수의 1/2이 되는 지점을 기본주파수라 잘못 결정하게 된 경우 이에 해당하는 배음행렬의 행은 [1 0 1 0 1 0 1 ...]과 같은 형태를 띄게 되어 기본주파수가 아님을 나타내게 된다. 또한 1/3인 지점을 기본주파수라 잘못 결정하게 된 경우에도 행렬은 [1 0 1 0 0 0 1 0 0 ...]같은 형태를 띄게 되어 기본주파수가 아님을 알 수 있도록 해준다. 결과적으로 이러한 성질을 이용하여 옥타브 에러를 검색하여 수정할 수 있게 된다, 이때, 최고 가중치에 해당하는 주파수가 옥타브 에러로 판명될 경우 해당 가중치 값을 0으로 할당한 후 다시 최고 가중치 값을 검색하고 이에 해당하는 배음행렬의 형태를 통해 다시 기본주파수로서의 결정이 합당한지를 검증하게 된다. 이러한 배음행렬의 도입을 통해 기존의 P&G 방법과 SHS 방법에서의 가장 빈번한 오류인 옥타브 에러를 줄일 수 있게 된다.

III. 실험결과

제안된 방법을 여러 악기음에 적용한 결과를 그림3~그림5에 나타내었다. 각각의 악기음에 대하여 SHS 방법, P&G 방법과 제안된 방법순으로 나열하여 비교하였다. 악기음의 소스로는 44.1kHz, 16bit로 샘플링된 악기음 연주록을 사용하였다. FFT 포인트는 2048로 하였으며 한 프레임의 길이는 40ms이다.

그림에서 알 수 있듯이 제안된 방법의 경우 SHS와 P&G방법 모두에 있어 보다 나은 음고 추정 결과를 나타내고 있다. 또한 SHS와 P&G 방법의 경우 음고가 변하는 부분에서의 불연속을 나타내고 있지만 제안된 방법에서는 음고가 변하는 부분에서 또한 만족스러운 만한 결과를 나타내고 있다.

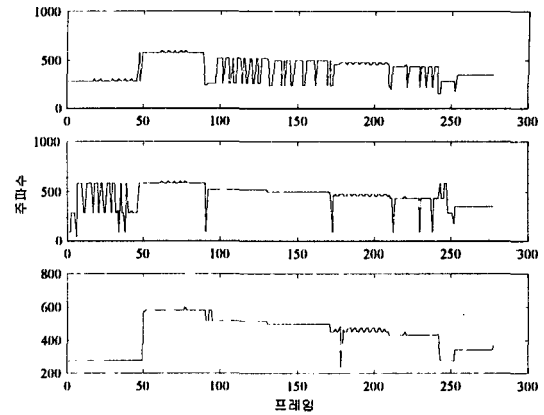


그림 3. 플루트 연주음에 대한 음고 추정결과

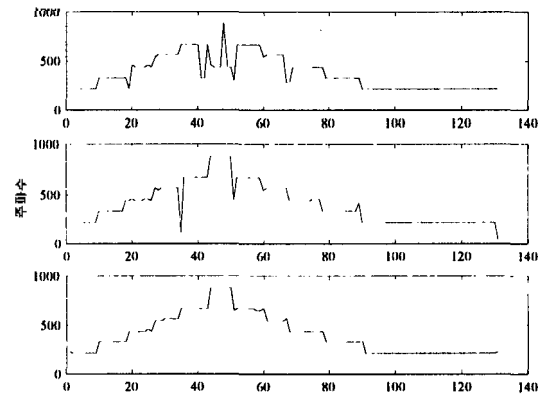


그림 4. 트럼펫 연주음에 대한 음고 추정결과

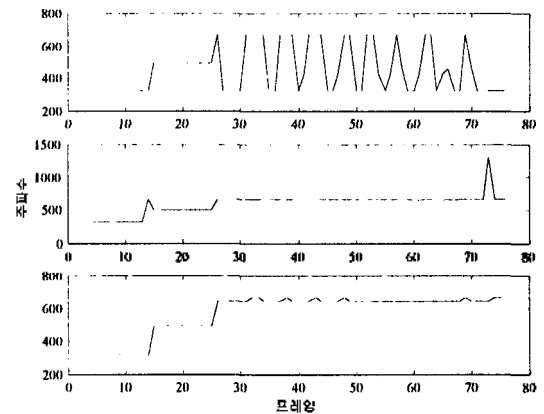


그림 5. 오보에 연주음에 대한 음고 추정결과

IV. 결론

기존의 SHS 방법과 P&G 방법을 개선하고 조합하여 자동음고 추정에 있어 오류 발생의 비율을 줄일 수 있는 방법을 제안하였다. 특히, 잡음 성분에 의한 오류의 발생 비율을 줄일 수 있었으며, 가장 빈번히 발생하는 오류인 옥타브 에러의 비율 또한 줄일 수 있었다.

향후 제안한 방법을 좀더 개선하여 오류를 더욱 줄일 수 있도록 하는 연구가 필요하며 또한, 복수의 악기음이 존재하는 경우에 있어 각각의 기본주파수를 추정하는 방법에 관한 연구가 필요하다.

참고문헌

- [1] M. Piszczalski, B. A. Galler, "Predicting musical pitch from component frequency ratios," J. Acoust. Soc. Am., Vol. 66, No. 3, pp. 710-720, 1979
- [2] R.C. Maher, J.W. Beauchamp, "Fundamental frequency estimation of musical signals using a two-way mismatch procedure," J. Acoust. Soc. Am., Vol. 95, No. 4, pp. 2254-2263, 1994
- [3] D. J. Hermes, "Measurement of pitch by subharmonic summation" J. Acoust. Soc. Am., Vol. 83, No. 1, pp. 257-264, 1988
- [4] J. C. Brown, "Musical fundamental frequency tracking using a Pattern recognition method," J. Acoust. Soc. Am., Vol. 92, No. 3, pp. 1394-1402, 1992