

진동수 패턴을 이용한 아기 울음소리 감지 방법*

박주건*, 임승현*, 윤준영*, 박기현*

*계명대학교 컴퓨터공학과

e-mail:corea@kmu.ac.kr

A Method for Detection of Baby Crying Using Frequency Pattern

Ju-Geon Pak*, Sung-Hyun Im*, Jun-Young Yoon*, Kee-Hyun Park*

*Dept of Computer Engineering, Keimyung University

요 약

최근 음성인식에 관한 연구가 활발히 진행되고 있지만, 음성으로 기기를 작동시키거나 신원을 파악하는 등과 관련된 성인 음성 인식에 관한 연구들이 대부분이고 아기의 울음소리를 감지하기 위한 학술적인 연구는 미비한 실정이다. 아기들은 통상 울음으로서 자신의 상태를 표현한다는 점을 고려해볼 때 기존의 성인을 대상으로 한 연구 결과를 그대로 적용시키기에는 무리가 따른다. 아기의 울음소리를 정확히 감지할 수 있다면 아기 및 유아를 위한 다양한 헬스(케어)기기에 적용될 수 있을 것이다. 따라서 대부분의 헬스(케어)기기들이 가지는 제한적인 자원과 컴퓨팅 능력을 고려하여 간단하면서도 정확도가 높은 방법이 필요하다. 이에 본 논문에서는 아기 울음소리의 진동수 패턴을 통계적으로 분석하여 아기 울음소리를 감지하는 방법을 제안한다. 다양한 주변 소리 샘플들을 통해 본 논문의 방법을 검증해본 결과 오감지율이 8.1%로 우수한 결과를 얻을 수 있었다.

1. 서론

최근 소득수준의 증가로 인해 건강의 중요성이 커지고 있고 IT 기술의 발전으로 인해 IT 용·복합 헬스기기 시장의 규모가 점점 커져 가고 있다. 이 중에서 최근 신생아 및 유아를 위한 상품들이 잇따라 출시되고 있다. 현재 신생아 및 유아의 대부분이 부모의 직장 생활 및 기타 이유로 인해 부모와 떨어져 있는 경우가 많다. 이렇다보니 아기들은 과거에 비해 정서 불안 및 스트레스 등을 많이 겪게 되고 언어를 통한 의사 전달 능력이 없는 아기들은 통상 울음으로서 자신의 상태를 나타낸다.

이러한 상황에서 아기의 울음을 그치게 하기 위한 다양한 제품들이 개발되고 있으며, 그 중 아기의 울음을 감지하여 산모의 심장박동음을 들려주는 등 정신적 안정 및 인성을 도와줄 수 있는 제품이 개발되고 있다. 실제 아기가 울 때 산모의 자궁에서 채취한 심장소리를 들려주는 제품이 개발되어 효능이 검증된 적이 있고 미국 병원에서는 이러한 제품을 사용 중이며 좋은 호응을 얻고 있다.

이러한 제품의 핵심은 아기의 울음소리를 정확히 감지할 수 있는 음성인식 기술에 있다. 하지만 현재까지 이와 관련된 학술적인 연구는 아직 미비한 실정이다. 대부분의 음성인식 기술은 각 단어 또는 모음과 자음을 발음할 때 발생하는 파형의 주파수 패턴을 통계적으로 분석하여 다

양한 소리의 주파수 패턴과의 비교를 통해 음성을 인식하는 방법을 사용하고 있다[1-4]. 하지만 아기들은 통상 울음으로서 자신의 상태를 나타낸다는 점을 고려해볼 때 기존의 성인을 대상으로 한 음성 인식 연구 결과를 그대로 적용시키기에는 무리가 따른다.

이에 본 논문에서는 본 논문의 방법이 적용될 헬스(케어) 제품의 제한된 자원과 컴퓨팅 능력을 고려하여 비교적 간단하면서도 정확도가 높은 아기 울음소리 감지 방법을 제안한다. 이를 위해 다수의 아기 울음소리 샘플을 통해 울음상태일 때의 특성을 파악하였다. 아기의 울음소리는 대부분 고주파(고음)로 단위 시간 당 진동수가 타 주변 소리에 비해 많았으며 중간 중간에 훌쩍거리거나 숨고르기를 함으로 인해 진동수가 급격히 감소하는 패턴을 가지고 있다. 이러한 특성을 고려하여 마이크로부터 입력되는 소리의 진동수를 분석하여 울음 후보 영역을 도출한다. 아기의 울음소리 중 중간에 진동수가 급격히 감소한다는 특징을 고려하여 도출된 후보 영역의 지속시간을 분석하여 최종적으로 아기 울음상태를 감지하는 방법을 제안한다.

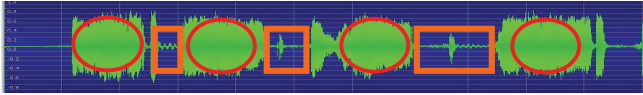
본 논문의 아기 울음소리 감지 방법을 총 37개의 다양한 소리에 적용해 본 결과 오감지율은 전체 37개 샘플 중 3개인 8.1%로 좋은 결과를 보여주었다.

2. 아기 울음 감지

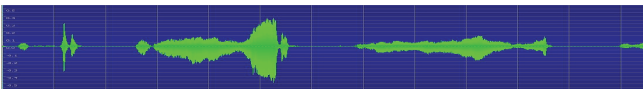
본 논문에서는 아기 울음소리의 패턴을 통계적으로 분석하여 특정 소리가 발생했을 때 패턴 매칭 방법으로 아

* 본 과제(결과물)는 교육과학기술부의 재원으로 한국연구재단의 지원을 받아 수행된 광역경제권 선도산업 인재양성사업의 연구결과입니다.

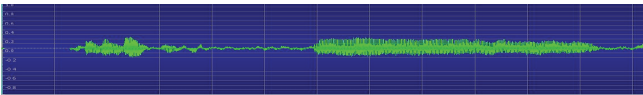
기 울음소리를 감지하고자 한다. 이를 위해서 소리 파형의 특징을 나타내는 대표적인 값인 진동수와 진폭을 이용하여 주변의 다양한 소리의 패턴을 분석하였다[5]. 그림 1과 2는 아기의 울음소리 파형을, 그림 3은 성인 남성의 대화 소리 파형을 나타낸다. 각 파형은 6초 동안 마이크를 통해 연속적으로 입력된 소리의 파형이다.



(그림 1) 아기 울음소리 파형 #1 - 마이크와 가까운 상황



(그림 2) 아기 울음소리 파형 #2 - 마이크와 떨어진 상황



(그림 3) 성인 남성의 대화 소리

그림 1과 3을 비교해보면 두 파형에서의 진동수와 진폭의 차이는 육안으로도 확연히 차이가 남을 알 수 있으며 그림 1의 파형이 그림 3의 파형에 비해 대체적으로 진동수가 많으며 진폭이 크다는 것을 알 수 있다. 이는 곧 아기의 울음소리는 다른 소리에 비해 대체적으로 크며 고주파(고음)임을 나타낸다. 이 외 기타 주변의 소리에서도 유사한 차이가 발생하였다. 따라서 아기 울음소리를 구분할 수 있는 파형의 특징으로 진동수와 진폭을 사용할 수 있을 것이다. 하지만 그림 1과 그림 2의 파형을 보면 둘 다 아기 울음소리에 관한 파형으로 비슷한 진동수를 보이지만 그림 1의 파형은 진폭이 큰데 반해 그림 2의 진폭은 비교적 작다. 이는 아기 울음소리를 입력받는 마이크와 아기 간의 거리에 의해 진폭의 값이 달라지기 때문이다. 마이크와 아기 간의 거리가 가까울수록 진폭은 커지며 멀수록 진폭은 작아지므로 같은 소리라도 거리에 따라 다른 특징 값을 보일 수 있다. 따라서 파형의 특성 중 진폭은 본 논문에서는 고려하지 않고 진동수를 이용하여 아기의 울음소리를 감지하도록 하였다.

아기 울음소리 파형의 또 다른 특징으로는 진동수의 변화를 들 수 있다. 그림 1에서 원으로 표시된 영역은 진동수가 많은 영역을 나타내며 사각형으로 표시된 영역은 진동수가 적은 영역을 나타낸다. 그림 1에 따르면 연속적으로 입력되는 아기 울음소리의 파형은 진동수가 많아지다가 급격히 감소하는 패턴을 반복한다. 이는 아기가 우는 중간 중간에 훌쩍거리거나 숨고르기를 하기 때문이다. 따라서 진동수가 많은 영역의 지속시간도 고려하여야 한다. 이에 이러한 아기 울음소리의 패턴을 이용하여 마이크로부터 입력되는 소리에서 진동수를 감지하여 아기 울음 후

보 영역을 결정하고 후보 영역이 지속되는 시간을 분석하여 최종적으로 아기 울음을 감지한다.

2.1 진동수를 이용한 울음 후보 영역 결정

현재 아날로그 음파를 디지털 신호로 변화하여 사용하는 대부분의 기기들은 일정 주기로 아날로그 음파의 진폭을 측정된 뒤(sampling), 이를 디지털 신호로 변화시키기 위해 PCM (Pulse Code Modulation) 방식을 사용하고 있으며 1초당 8000번 샘플링하고, 각 샘플링된 데이터를 8비트로 표현하고 있다[5]. 이에 본 논문에서도 1초당 8000번 샘플링된 8비트의 값을 기준으로 하여 아기의 울음소리를 분석한다. 다양한 아기의 울음소리를 통계적으로 분석한 결과 평균적으로 매 0.1초 구간동안 90번 이상의 진동이 있었음을 알 수 있었다. 이러한 실험적 결과를 임계값 $T(=90)$ 로 하여 울음 후보 영역을 결정한다.

0.1초 동안 샘플링된 값 중 n 번째 샘플링된 값을 X_n 라고 했을 때, X_n 의 바로 이전 샘플링된 값은 X_{n-1} 이 된다. 0.1초 동안의 진동수를 계산하기 위해 파형에서 부호가 변경된 횟수 $sign_count$ 를 카운트하고 누적된 $sign_count$ 를 이용하여 매 0.1초 구간 N 에서의 진동수 $F[N]$ 을 계산한다.

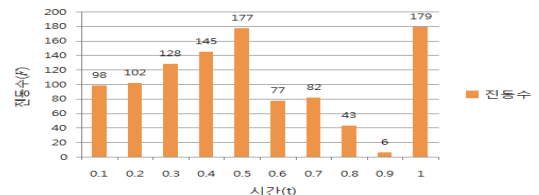
$$\begin{cases} \text{if } (x_n \times x_{n-1}) < 0 : & sign_count++; \quad ; n = 1, 2, \dots \\ \text{else} & : no-op; \end{cases} \quad (1)$$

$$F[N] = \lfloor sign_count / 2 \rfloor \quad ; N = 1, 2, \dots$$

N 번째 0.1초 구간의 진동수 $F[N]$ 이 계산되면 조건식 (2)를 사용하여 진동수 $F[N]$ 이 임계값 T 이상이면 울음 후보 영역 $Candidate_Region[N]$ 을 1로, 그렇지 않다면 0으로 설정한다.

$$\begin{cases} \text{if } F[N] \geq T : & Candidate_Region[N] = 1; \\ \text{else} & : Candidate_Region[N] = 0; \end{cases} \quad (2)$$

예를 들어 아기의 울음소리가 2초 간 마이크로부터 입력되었을 때, 식 (1)를 사용하여 매 0.1초 구간별로 진동수를 계산하면 그림 4와 같이 나타낼 수 있다. 계산된 진동수는 조건식 (2)에 의해 그림 5와 같이 울음 후보 영역 $Candidate_Region[N]$ ($N=1, 2, 3, \dots, 10$) 으로 설정할 수 있다.



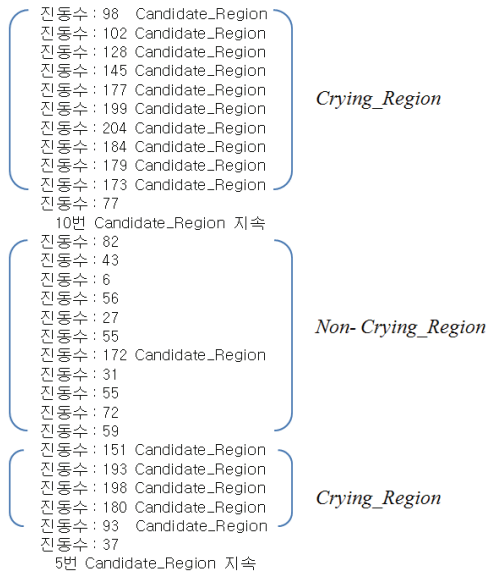
(그림 4) 입력 소리의 진동수 그래프

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| [1] | [2] | [3] | [4] | [5] | [6] | [7] | [8] | [9] | [10] |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |

(그림 5) $Candidate_Region$ 의 값

2.2 지속시간을 고려한 울음 감지

마이크를 통해 연속적으로 입력되는 아기 울음소리의 파형은 진동수가 많아지다가 급격히 감소하는 패턴을 반복한다. 이는 아기가 우는 중간 중간에 훌쩍거리거나 숨고르기를 하기 때문이다. 따라서 진동수 계산을 통해 결정된 아기 울음 후보 영역 *Candidate_Region*의 지속시간도 고려해야 한다. 다양한 아기의 울음소리를 통계적으로 분석한 결과 *Candidate_Region*이 연속 3번(0.3초) 이상 나타남을 알 수 있었다. 또한 아기가 우는 중간 훌쩍거림 또는 숨고르기로 인해 *Candidate_Region*이 30번(3초) 이상은 연속되지 않음을 알 수 있었다. 즉 *Candidate_Region*이 3번(0.3초)에서 30번(3.0초) 정도는 지속적으로 감지되어야 울음 영역 *Crying_Region*이라고 판단할 수 있다. 그림 6은 *Crying_Region*을 결정 방법을 나타낸다.



(그림 5) *Crying_Region* 판단 과정

그림 6에서와 같이 마이크로부터 연속적으로 입력되는 소리는 식 (1)에 의해 진동수가 계산되고 식 (2)에 따라 *Candidate_Region*이 설정된다. *Candidate_Region*이 최소 3번(0.3초)에서 최대 30번(3.0초)까지 지속적으로 설정되면 그 구간 전체를 하나의 *Crying_Region*으로 결정한다.

*Crying_Region*이 결정되면 매 10초마다 누적된 *Candidate_Region*과 *Crying_Region*의 관계에 의해 최종적으로 아기 울음 여부를 판단한다. 실험적 결과에 의하면 *Candidate_Region*을 *Crying_Region*으로 나눈 값은 7~17임을 알 수 있었다. 이러한 통계적 결과를 토대로 조건식 (3)을 통해 최종적으로 아기 울음 여부를 감지하게 된다.

$$Candidate_Count = \sum_{N=1}^{100} Candidate_Region[N]$$

$$Crying_Count = \sum_{N=1}^{100} Crying_Region[N]$$

$$\begin{cases} \text{if } 7 \leq (Candidate_Count / Crying_Count) \leq 17 : Crying \\ \text{else} : Non-Crying \end{cases} \quad (3)$$

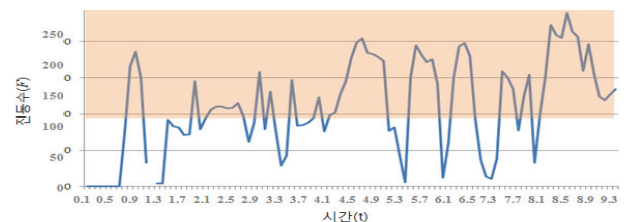
3. 실험 결과

제안한 방법의 성능을 평가하기 위해서 총 37개의 음성 샘플을 사용하였으며 그 결과를 표 1에 정리하였다. 표 1에서 7개의 서로 다른 아기 울음소리 샘플에 대해 감지율(Detection Rate)이 100%임을 알 수 있으며 미감지율(Missing Rate)이 0%임을 알 수 있다. 평균 오감지율(False Positive Rate)는 전체 37개 샘플 중 3개인 8.1%로 좋은 결과를 보여준다.

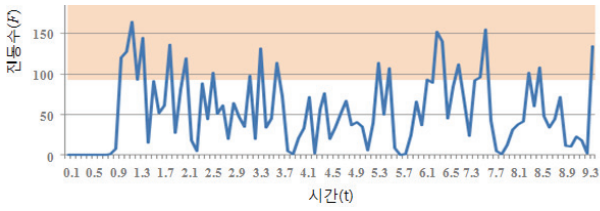
<표 1> 아기 울음소리 감지 결과

| 종류 | 감지 | 오감지 | 미감지 |
|-------------|-----------|----------|----------|
| 아기울음소리(7) | 7 | 0 | 0 |
| 남자 댄스노래(1) | 1 | 0 | 0 |
| 남자 발라드노래(1) | 1 | 0 | 0 |
| 여자 댄스노래(1) | 0 | 1 | 0 |
| 여자 발라드노래(1) | 1 | 0 | 0 |
| 강의 소리(2) | 2 | 0 | 0 |
| 남성 저음(2) | 2 | 0 | 0 |
| 남성 평음(2) | 2 | 0 | 0 |
| 남성 고음(2) | 2 | 0 | 0 |
| 남성 기타(1) | 1 | 0 | 0 |
| 여성 저음(2) | 2 | 0 | 0 |
| 여성 평음(2) | 2 | 0 | 0 |
| 여성 고음(2) | 2 | 0 | 0 |
| 여성 기타(1) | 1 | 0 | 0 |
| TV 방송 소리(4) | 3 | 1 | 0 |
| 광고소리(5) | 4 | 1 | 0 |
| 총소리(1) | 1 | 0 | 0 |
| 합계 | 34 | 3 | 0 |

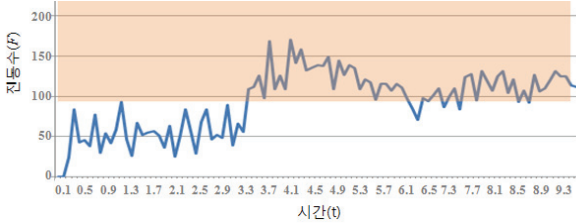
그림 6, 7, 8, 9, 10은 표 1의 샘플 중 아기 울음소리, TV 방송소리-뉴스, 여자 댄스노래, TV 방송소리-기타, 광고 소리-기타를 진동수 그래프로 나타낸 것이다.



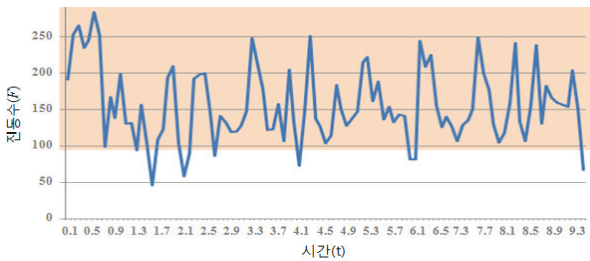
(그림 6) 아기 울음소리



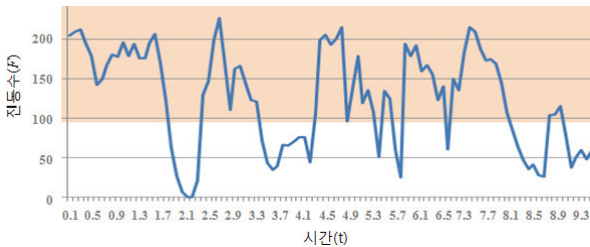
(그림 7) TV 방송 소리 - 뉴스



(그림 8) 여자 댄스 노래의 진동수 그래프



(그림 9) TV 방송 소리-기타



(그림 10) 광고 소리-기타

표 2는 진동수 그래프로 변환된 샘플에 대해 계산된 Candidate_Region과 Crying_Region을 나타낸다.

<표 2> 각 샘플의 Candidate_Region과 Crying_Region

| 종류 | Candidate_Region | Crying_Region | 계산 결과 | 감지 결과 |
|-------|------------------|---------------|-------|------------|
| 그림 6 | 22 | 2 | 11 | Crying |
| 그림 7 | 9 | 2 | 4.5 | Non-Crying |
| 그림 8 | 33 | 4 | 8 | Crying |
| 그림 9 | 42 | 3 | 14 | Crying |
| 그림 10 | 66 | 6 | 11 | Crying |

그림 6의 아기 울음소리는 계산 결과가 11로 Crying로 판단하였고 그림 7의 뉴스 소리는 4.5로 Non-Crying로 정

상적으로 판단하였다. 하지만 그림 8, 9, 10의 경우 아기 울음소리가 아님에도 불구하고 계산 결과에 의해 Crying으로 오감지하였음을 알 수 있다. 이는 오감지한 3개의 샘플은 다른 샘플들에 비해 고음이고, 고음이 지속되는 부분이 많아 진동수가 90 이상인 부분이 아기의 울음소리와 비슷한 패턴으로 나타나기 때문이다.

4. 결론

본 논문에서는 주변의 다양한 소리 중 아기의 울음소리를 감지할 수 있는 방법을 제안하였다. 아기의 울음소리 감지 기술은 아기가 울고 있을 때 산모의 심장 박동음을 들려주는 등 다양한 헬스(케어) 제품에 적용될 수 있다.

본 논문에서는 헬스(케어) 제품의 제한된 자원과 컴퓨팅 능력을 고려하여 비교적 간단하면서도 정확도가 높은 아기 울음소리 감지 방법을 제안하였다. 이를 위해 주변의 다양한 소리들의 특징을 분석하였고 아기 울음소리가 가지는 진동수 패턴과 진동수 지속시간을 고려하여 아기 울음소리를 감지할 수 있는 방법을 제안하였다. 제안된 방법의 성능 평가를 위해 총 37개의 음성 샘플을 사용하였으며 그 결과 아기 울음소리 샘플에 대해 감지율이 100%임을 알 수 있으며 미감지율이 0%임을 알 수 있다. 또한 평균 오감지율 (False Positive Rate)는 전체 37개 샘플 중 3개인 8.1%로 좋은 결과를 보여줌을 검증하였다.

하지만 실생활 환경에서 아기 울음소리 감지율을 높이기 위해 더 많은 샘플을 대상으로 한 실험이 추가적으로 필요하며 아기 울음소리가 아닌 기타 주변의 소리에 대한 오검출률(False Positive Rate)를 줄여야 할 필요성이 있다.

참고문헌

- [1] Lawrence R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", Proceedings of the IEEE, Vol. 16, Issue 2, pp. 257-286, 1989.
- [2] R. P. Lippmann, "Review of Neural Networks for Speech Recognition", Neural Computation, Vol. 1, Issue 1, pp. 1-38, 1989.
- [3] 지상문 "음성인식을 위한 주파수 부대역별 효과적인 특징추출", 한국해양정보통신학회 논문지, 제 7권, 제 5호, pp. 598-603, 2003.
- [4] 송원문, 김명원, "문맥 및 사용 패턴 정보를 이용한 음성인식의 성능 개선", 정보처리학회논문지(B), 제 13권, 제 5호, pp. 553-560, 2006
- [5] 유지상, 박호중, 심동규, "디지털 신호처리", 3판, 생능출판사.