음성인식률 향상을 위한 잡음 제거

황동환 / 엑스텔테크놀로지

개 요

많은 연구를 통해 음성 인식은 잡음이 존재하지 않는 환경에서는 매우 높은 인식률을 보이고 있으며 실제로 여러 분야에서 응용되고 있다. 하지만 여러 잡음이 존재하는 환경에서는 그 성능이 급격하게 저하되어 잡음에 둔감한 인식기와 잡음 제거가 필수적이다. 본 내용에서는 독립 요소 기법에 기반 한 잡음 제거 기법을 소개하고 이를 칩으로 구현하고 그 결과를 고찰해 보겠다.

독립 요소 해석 기법에 기반한 적응 잡음 제거

필터는 데이터로부터 유용한 정보를 추출하는 방법으로 여러 응용 분야에서 다양한 목적으로 사용되어 왔다. 이러한 필터를 설계함에 있어 다양한 방법이 존재하지만, 필터의 출력과 원하는 출력의 평균 제곱을 최소화시키는 방법인 Wiener 필터는 적응 잡음 제거, 적응 반향 제거, 적응 선형 예측 등에 적응 신호 처리 시스템에 많이 사용되고 있다. 그 중 본 연구의 관심 분야인 적응 잡음 제거 시스템을 그림 2.1에 나타내었다.

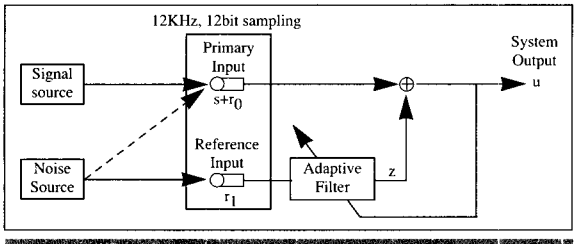


그림 2.1 적응 잡음 제거 시스템 블록도

여기서 일반적인 학습 방법은 u=s+r_0+-z출력 의 제곱 평균을 최소화 하는 해를 찾는 것이고 최소 평균 제곱을 갖도록 적응 필터를 학습해 가는 것을 기반으로 하고 있으며, 신호 s, r_0, r_1, z가 통계적으로 정상(statistically stationary)이고, s는 r_0, r_1과 서로 비상관(uncorrelated)이고 r_0, r_1은 서로 상관(correlated)이라고 가정한다.

독립 요소 해석

독립 요소 해석은 알지 못하는 채널을 통해 섞인 음원들의 혼합 신호를 입력으로 받아들여 상호 독립적인 음원 신호를 복원해 내는 기술이다. 알지 못하는 신호 $s(n)$=$[s_1(n), s_2(n) \cdots, s_N(n)]$의 평균이 0이고 상호 독립이라고 가정하자. 개의 센서로부터 미지의 신호원의 선형 혼합 신호 $x(n)$=$[x_1(n), x_2(n),\cdots, x_N(n)]^T$가 관찰되어진다고 할 때, 가정에 의해 $x(n)$은 다음 식으로 표현될 수 있다.

$$x(n) = A \times s(n)$$

행렬 는 미지의 가역 행렬로 혼합 행렬이라고 부른다. 여기서 측정 가능한 만을 이용하여 혼합 행렬 의 역행렬을 찾음으로써 다음과 같이 음원 신호 을 얻는 것이다.

$$s(n) = A^{-1} \times x(n)$$

그러나, $x(n)$만을 이용하여 정확한 A^1을 추정할 수 없기 때문에, 이를 해결하기 위해서 다음 두 가지를 허용해 준다. 첫째는 각 음원 신호의 순서를 입력에 들어온 그대로 복원할 수 없기 때문에 복원한 신호는 순서가 뒤바뀐 음원 신호를 얻을 수 있다는 것이며, 둘째는

복원한 신호가 원래 음원 신호의 크기를 보존할 필요가 없다는 것이다. 이것은 신호의 분리 측면에서 심각한 문제를 일으키지 않으며, 이러한 조건하에 추정해야 할 분리 행렬 는 다음과 같은 식으로 나타낼 수 있다.

$$W=P \cdot A^{-1}$$

분리 행렬 W를 추정하기 위해 음원들이 서로 독립이라는 가정을 설정한다. 이는 하나의 음원에서 나오는 신호가 다른 음원에서 나오는 신호에 영향을 주지 않는다는 것을 의미하는 것으로 실세계에서 충분히 개연성 있는 가정이라고 할 수 있다. 그리고, 통계적 독립성이 크기와 순서에 무관하기 때문에 분리 행력을 찾게 된다. 이렇게 추정한 분리행렬을 이용하여 다음과 같은 출력 신호를 얻게 된다.

$$u(n)=W \cdot x(n)$$

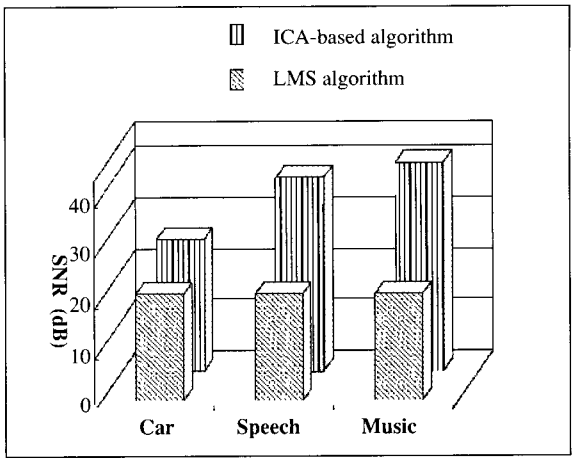
Bell과 Sejnowski는 분리 행렬 W를 학습하는 방법으로 입력 신호의 누적 확률 분포 g을 통과한 출력 y=g(u)의 상호 정보(mutual information)라는 비용 함수를 제시하였다.

Bell과 Sejnowski는 음성 신호를 포함한 많은 실세계의 신호가 super-Gaussian 분포를 따르게 되는데, 이 때에는 출력 신호 의 결합 엔트로피를 최대화함으로써 출력 신호 간의 상호 정보를 최소로 할 수 있음을 지적하였으며, 다음과 같은 수식으로 나타낼 수 있다.

$$H(y)=-E[\ln(p(y))]$$

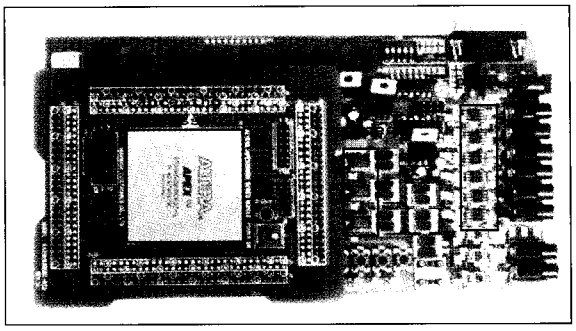
독립 요소 해석 기법을 이용한 적응 잡음 제거

LMS(최소 평균 제곱) 학습 법칙과 ICA(독립 요소 해석 기법) 학습 법칙을 비교하면 가장 큰 차이점은 비용함수이다. LMS 학습 법칙은 출력과 상관 관계가 있는 참조 잡음(referenced noise)을 이용하여 잡음을 제거하는 반면 ICA 학습 법칙은 고차 통계학적인 의미를 갖는 상호 독립인 신호로 분리한다. 이는 잡음원이 통계학 적으로 서로 의존적인 다른 경로를 통하여 유입되는 경우 ICA 학습 법칙이 LMS 방법에 비해 적응 잡음에 대해 더 나은 성능을 보이며, 비용 함수를 $sign(\cdot)$로 사용하였을 때 학습 규칙이 곱하기 연산 대신 부호 연산으로 간단하게 바뀌면서도 더 나은 성능을 보이는 것을 알 수 있다. 그림 2.2에 신호원을 음성 신호로 하고 잡음원을 달리했을 때 인식률 향상을 나타내었다. 독립 요소 해석 기법에 의한 방법이 LMS 학습 법칙에 비하여 학습률이 간단하면서도 더 나은 성능을 보이고 있다.



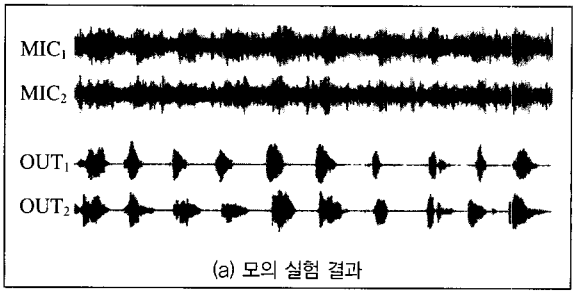
FPGA를 이용한 실세계 적용

모의 실험 결과 본 논문에서 제안된 잡음 제거 프로세서의 실생활 적용 가능성을 시험하기 위하여 Field Programmable Gate Array(FPGA)를 이용하였다. Altera Flex20K600EBC-1 FPGA 칩을 이용하여 잡음 제거 프로세서를 이용하였으며 입력 신호 6개, 출력 신호 6개를 출력할 수 있는 보드를 작성 하였다. 입력 신호는 모노 신호이고 출력 신호는 스테레오 신호이다. A/D, D/A Codec은 PCM3002 칩을 이용하였고 12 KHz, 12 bit로 샘플링 하였다. 이를 그림 2.3에 나타내

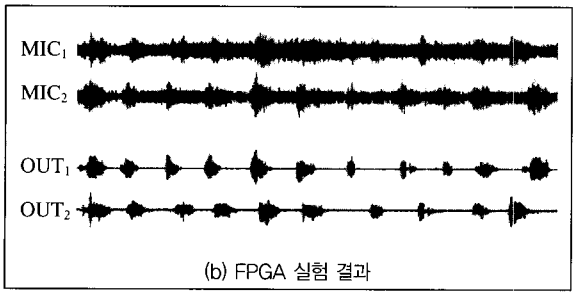


었다. 작성된 시험 보드를 테스트 하기 위하여 5개의 스피커를 지원하는 파워 앰프를 사용하였고, 각 잡음원은 MP3 Player를 이용하였다. 시험 보드는 6개의 입력 신호를 받아 들일 수 있도록 설계되었기 때문에, ICA 음원을 2개, ANC 음원을 4개로 하였다. 이는 마이크를 2개 사용하였다는 것을 의미한다.

첫번째 실험에서 ICA 음원은 2개의 음성으로 이루어져 있고, ANC 음원은 4개의 서로 다른 음악 신호로 구성되어 있다. 그리고 두번째 실험은 ICA 음원을 실제 사람의 음성 신호와 스피커로부터 출력되는 자동차 잡음으로 이루어져 있고, ANC 잡음원은 4개의 음악으로 이루어져 있다.



(a) 모의 실험 결과



(b) FPGA 실험 결과

그림 2.4 잡음 제거 프로세서 성능 시험:2 ICA[영어 숫자음, 남성(스피커), 여성(스피커)]+4 ANC[음악]]

첫번째 실험은 추후에 소개될 음성 인식 시스템과의 융합의 타당성을 보이기 위하여 실제로 인식 실험에 사용하였던 영어 숫자음의 남성, 여성 음성을 ICA 음원으로 사용하였다. 또한 ICA 음원은 스피커를 통하여 인가하였다. 모의 실험 결과와 실세계 실험 결과를 그림 2.4에 나타내었다. FPGA를 이용한 실험 결과가 모의 실험 결과보다 약간 성능이 저하되지만 거의 유사한

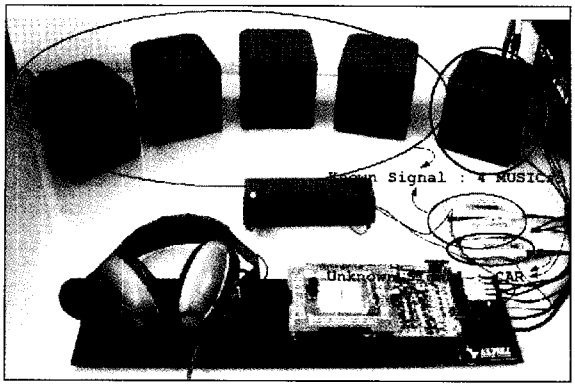


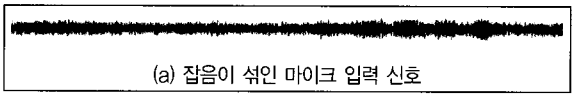
그림 2.5 잡음 제거 프로세서 성능 실험을 위한 시스템 구성

결과를 보이고 있어 모의 실험에서 가정한 실험 환경이 실세계와 유사함을 알 수 있다.

가능성을 좀 더 실세계에 가까운 실험을 하기 위하여 ICA 음원을 실제 사람의 음성과 자동차 잡음을 스피커로 인가하여 실험하였다.

이를 그림 2.5에 나타내었고 그림 2.6에 그 결과를 나타내었다.

모의 실험과는 달리 신호원을 알 수 없기 때문에 직접적인 SNR 향상 정도를 구하기는 힘들지만, 묵음 구간의 잡음 제거 성능으로 SNR 향상을 유추해 볼 수 있


(a) 잡음이 섞인 마이크 입력 신호

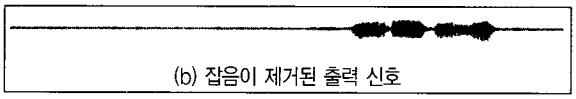

(b) 잡음이 제거된 출력 신호

그림 2.6 실세계 잡음 제거 프로세서 성능 실험:2 ICA[음성(육성), 자동차 잡음(스피커)]+4 ANC[음악]]

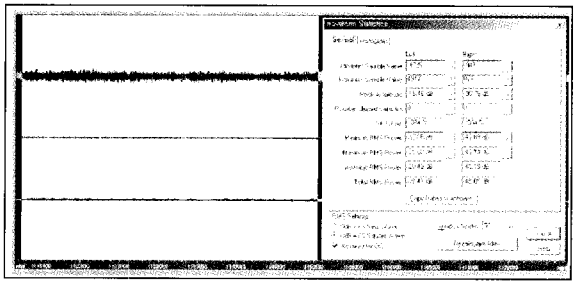


그림 2.7 실세계 잡음 제거 프로세서 실험 결과:SNR 향상

다. 이를 그림 2.7에 나타내었다.

FPGA를 이용한 실세계 잡음 제거 실험에서도 SNR 향상이 16.7 dB 정도 향상되었음을 알 수 있으며, 잡음 제거 프로세서의 실세계 응용 가능성이 매우 높은 것을 보여주고 있다.

결 론

실세계에는 여러 가지 형태의 잡음이 존재하고 이는 음성 인식의 성능을 크게 저하시키고 있다. 본 논문에서는 일반적으로 실세계에 존재하는 잡음이 서로 독립이라는 점과 상호 독립인 신호를 분리시키는 독립 요소 기법(ICA)을 이용하여 잡음을 효과적으로 분리하는 잡음 제거 프로세서를 제안하였다. 실세계에서는 신호원이 다양한 경로를 거쳐 센서에 입력되므로 이를 역상관하여 제거해야만 하므로 그 계산량이 많아 범용 프로세서(Pentium4 2G)로는 3개 이상의 잡음원을 제거하기 힘들다.

본 논문에서는 모듈 개념을 도입하여 확장을 통해 16개의 잡음원을 알고 있는 잡음과 3개의 잡음원을 알지 못하는 잡음을 제거할 수 있는 잡음 제거 프로세서를 설계하였다.

FPGA를 이용하여 잡음원을 알고 있는 4개의 신호와 잡음원을 알 수 없는 2개의 신호가 섞인 환경에서 음성만을 추출하는데 성공하였으며 SNR이 약 16.7 dB 개선되었다. 하지만 여전히 잡음원이 점음원(point source)이여야 한다는 단점이 남아있고 센서의 위치나 잡음원의 위치가 바뀔때 재빨리 학습해야 하는 과제가 남아 있다. 하지만 대부분의 잡음원은 위치가 쉽게 바뀌지 않기 때문에 많은 분야에 적용할 수 있을 것으로 기대된다.

본 논문에서 제안된 잡음 제거 프로세서는 다양한 실세계 환경에서 실시간으로 잡음을 제거할 수 있으며 이를 이용하여 음성 인식이나 기타 잡음 제거가 필요한 시스템에 유용하게 쓰일 것으로 예상된다.