2 스피커 입체음향 재생용 크로스톡 제거기 유저 인터페이스

*김 현석, *김 풍민, *서 상원, *김 기홍

소속: *한국전자통신연구원 휴먼컴퓨팅연구부 청각정보연구팀

대전 유성구 가정동 161 번지 한국전자통신연구원 휴먼컴퓨팅연구부 청각정보연구팀

연락처: Email: s_kimm@audiows1.etri.re.kr

Cross-talk Canceler User Interface for Two-speaker 3D Sound Reproduction

Auditory Information Team, Human Computing Department, ETRI, Science Town, Taejon 305-701, KOREA

TEL: 82-42-860-1152, FAX: 82-42-860-5010

Email: s_kimm@audiows1.etri.re.kr

요약

스테레오로 생성된 음향을 헤드폰을 사용하면 좌, 우의 소리는 각각 좌,우의 귀로 상호 혼선 없이 전달된다. 그러나 2 개의 스피커를 사용한다면 좌,우의 소리가 혼합되어 들린다. 이러한 혼합현상을 크로스톡이라 하며, 이 경우 본래의 소리는 변질되어 들린다. 이러한 현상을 방지하기 위해 스피커로 재생하기 전 신호를 크로스톡 제거기를 통해 한번 더 처리하는 방법을 사용한다. 여기에 사용되는 크로스톡 제거기는 2 개의 스피커와 청취자간의 상대 위치, 그리고 스피커와 청취자가 있는 방의 음향특성에 의해서 적절히 변화되어야 한다. 본 연구에서는 방의 음향특성이 무향실과 같다고 가정하고 2 개의 스피커와 청취자간의 상대적인 위치가 주어졌을 때 효과적으로 크로스톡 제거기를 생성해주는 알고리듬을 제시하였다. 제안된 방법은 머리전달함수를 기반으로 하여 스피커가 청취자로부터 표준 반경에 위치하였을 때의 크로스톡 제거기를 데이터 베이스화 하여 일반적인 거리에 위치하였을 경우에도 실시간으로 확장 적용할 수 있다.

서론

입체음향에 관련된 기술은 생성, 녹음, 재생 등 여러 분야가 있다.[1] 그 중 크로스톡 제거기 구성 기술은 스피커를 이용한 재생분야에 속한다. 헤드폰을 위해 만들어진 입체음향 또는 스테레오 음향은 두개의 스피커를 통해 재생되면 두 스피커와 청취자의 위치에 따라서, 또한 스피커와 청취자가 있는 음장의 특성에 따라서 청자가 두 귀를 통해 듣게 되는 소리는 원음과 매우 달라질 수 있다. 그것은 오른쪽 스피커에서 나온 소리가 오른쪽 귀로만 전달되지않고 왼쪽으로도 전달되며, 마찬가지로 왼쪽 스피커에서 나온 소리가 왼쪽 귀로만 전달되지 않고 오른쪽으로도 전달되기 때문이다. 오른쪽 채널의 음이 왼쪽귀로, 왼쪽 채널의 음이 오른쪽 귀로 전달는 것을 크로스톡이라 하며, 이러한 크로스톡 현상은 스테레오 음향을 두 개의 스피커로 직접 내보내기 전 적당히 처리함으로써 제거될 수 있다[2,3].

그림 1 에서와 같이 크로스톡 제거기 $Q(z)$는 디지털 신호화 된 원 스테레오음향 신호와 스피커 구동신호 사이에 위치하게 되며, 2 입력 2 출력 시스템이 된다. 크로스톡 제거기를 통과한 두 신호는 역시 2 입력 2 출력 시스템인 스피커-청자간 전달함수 $H(z)$를 통해서 청

자의 양쪽 귀까지 도달하게 된다. 결국 크로스톡 제거기가 목표하는 것은 제거기 $Q(z)$와 재생계 $H(z)$의 곱이 항등함수 I 가 되게 하는 것으로 볼 수 있다. 그러나 $H(z)$가 비최소위상 함수인 경우엔 (일반적인 음향 시스템이 그러하다.) $Q(z)$가 실시간 구현이 불가능한 형태로 나타나므로 $H(z)Q(z)$가 적당한 지연시스템 $z^{-d}I$ 가 되도록 $Q(z)$를 만든다. 즉, 크로스톡 제거기 $Q(z)$의 형태는 $H(z)$의 역지연(Delayed-inverse)이 된다.

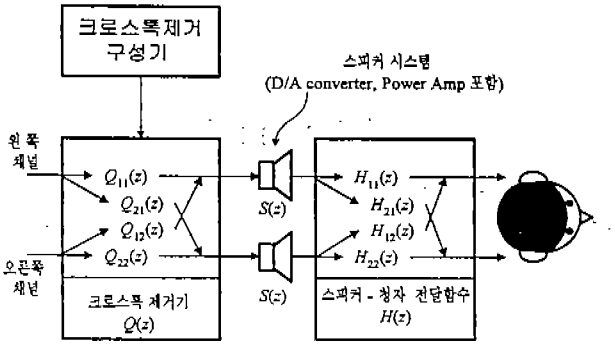

그림 1. 크로스톡 제거 시스템

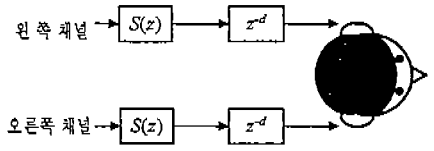

그림 2. 크로스톡 제거 후 스테레오음-청자간 전달계

2 입력 2 출력 전달함수 $H(z)$의 역지연 행렬을 적응필터를 사용하여 구할 수 있으며 일반적으로 Multiple Filtered-X LMS 알고리듬[2,3,4]을 이용한다. 이 방법은 $Q(z)$를 4 개의 FIR 적응필터로 구성하여 $H(z)Q(z)$ 와 Iz^{-d}의 차이가 Least Mean Squares 원리에 의해서 최소화되도록 $Q(z)$를 적응시킨다. $H(z)$가 바뀌면 다시 $Q(z)$를 적응시켜야 하며, 실제 스피커와 청자의 상대적인 위치, 방향정보가 변하면 재 계산이 필요하다. 크로스톡 제거기가 첨가됨으로써 2 채널 신호로부터 청자의 양쪽 귀까지의 전달함수는 그림 2 와 같이 스피커 재생부 $S(z)$를 거치는 단순 지연 시스템으로 변환되고 입체음향의 정보를 최대한 유지하여 청자에게 전달할 수 있다. 따라서 본 논문의 크로스톡 제거기 구성 접근 방법에서는 스피커 부분의 등화는 고려하지 않는다.

본 논문에서는 범용적인 크로스톡 제거기 유저인터페이스를 구성하기위해서 우선 $H(z)$가 머리전달함수(HRTF)를 기반으로 표현되도록 간략화 하였다. 따라서 스피커와 청취자가 위치해있는 음장은 무향실과 같다고 가정한다. 이러한 가정은 청취자가 스피커의 direct field 영향을 많이 받으며 방의 흡음률이 아주 낮지 않을 경우 실제와 그리 틀리지않는다.

효율적인 크로스톡 제거기 구성을 위해 다음과 같은 방법으로 접근하였다. 우선 스피커가 청취자로부터 표준반경 위에 있다고 생각하여 여러 좌 우 스피커 각도들에 대한 크로스톡 제거기 DB 를 적응 필터 기법을 이용하여 먼저 구성한 후 일반적인 스피커-청자 위치에 대하여는 적응과정 없이 간단한 맵핑을 통해 즉각적으로 크로스톡 제거기를 구성한다.

본 논문에서 제안된 크로스톡 제거기 구성 방법은 유저 인터페이스와 함께 모듈화 되었으며 주관적 평가 실험을 통해 매우 만족할만한 성능을 보여주었다.

이론

가. 스피커-청자모델

그림 3 과 같이 표준반경 r_0위에 두 스피커가 청자의 정면 방향으로부터 θ_1과 θ_2의 각도로 놓여있다고 가정하면 다음과 같이 4 개의 스피커-청자간 전달함수를 정의할 수 있다.

$H_L(\theta_1,z)$: 스피커 1과 좌측 귀 사이의 HRTF (1-a)

$H_R(\theta_1,z)$: 스피커 1과 우측 귀 사이의 HRTF (1-b)

$H_L(\theta_2,z)$: 스피커 2와 좌측 귀 사이의 HRTF (1-c)

$H_R(\theta_2,z)$: 스피커 2와 우측 귀 사이의 HRTF (1-d)

여기에서 아래 첨자 R 은 오른쪽, L 은 왼쪽을 의미한다. 그러면 그림 4 에서와 같은 일반위치 r_1, r_2거리에 있는 스피커와 청자간 전달함수는 다음과 같이 머리전달함수로만 표현된 부분인 $T(z)$와 r_0에 대한 거리변화를 표현하는 $M(z)$의 곱으로 표현된다.

$$
\begin{aligned}
H(z) &= \begin{bmatrix} H_{11}(z) & H_{12}(z) \\ H_{21}(z) & H_{22}(z) \end{bmatrix} \\
&= \begin{bmatrix} H_L(\theta_1,z) & H_L(\theta_2,z) \\ H_R(\theta_1,z) & H_R(\theta_2,z) \end{bmatrix} \begin{bmatrix} m(r_1,r_0) & 0 \\ 0 & m(r_2,r_0) \end{bmatrix} \\
&= T(z)M(z)
\end{aligned}
\qquad (2)
$$

여기에서 $m(r_1,r_0)$과 $m(r_2,r_0)$는 다음과 같이 정의된 양으로서 각각 r_1, r_2가 표준반경 r_0으로부터 벗어난

데 대한 스피커-청자간 전달함수의 이득변화와 시간지연변화를 보정하는 역할을 한다. (c 는 음속, f는 음향신호의 샘플링 주파수)

$$m(r_1,r_0)=\frac{r_0}{r_1}z^{-(r_1-r_0)f/c} \tag{3-a}$$

$$m(r_2,r_0)=\frac{r_0}{r_2}z^{-(r_2-r_0)f/c} \tag{3-b}$$

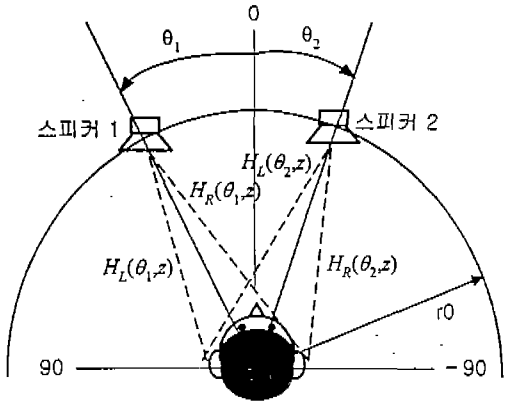

그림 3. 표준 반경 재생 시스템과 머리전달함수

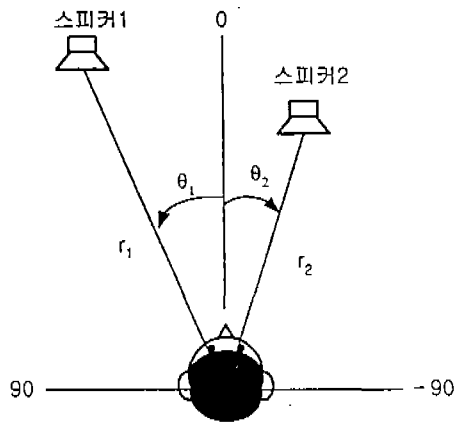

그림 4. 일반 위치 재생 시스템

나. 역지연함수와 고정반경 재생시스템의 데이터베이스화

식 (2)에서 스피커-청자간 2 행 2 열 전달함수 $H(z)$를 표준반경 r_0에서의 머리전달함수로 표현된 부분 $T(z)$와 거리와 관계된 함수 $M(z)$의 곱으로 표현하였다. 지연양을 d라 하고 $H(z)$의 역지연행렬을 $Q(z)$라 하면 다음이 성립한다.

$$H(z)Q(z)=z^{-d}I \tag{4}$$

식 (2)를 이용하면 $H(z)$의 역지연행렬 $Q(z)$는 $M(z)$의 역행렬과 $T(z)$의 역지연행렬인 $C(z)$ 의 곱으로 나타난다.

$$\begin{aligned}Q(z)&=M(z)^{-1}T(z)^{-1}z^{-d}\\&=M(z)^{-1}C(z)\end{aligned} \tag{5}$$

거리에 관련된 함수 $M(z)$ 의 역행렬은 다음과 같이 간단히 얻어진다.

$$M(z)^{-1}=\begin{bmatrix}m(r_0,r_1) & 0\\ 0 & m(r_0,r_2)\end{bmatrix} \tag{6}$$

따라서 $Q(z)$를 데이터 베이스화 하는 것보다 $C(z)$ 만을 데이터 베이스화 하는 것이 효율적이며, 이 점이 본 논문의 핵심적인 아이디어이다. 즉 일반위치의 데이터 베이스보다는 고정반경의 데이터베이스를 구축하여 이를 이용, $M(z)^{-1}$를 곱함으로써 일반위치의 크로스톡 제거기를 즉각 구성할 수 있다.

$C(z)$ 데이터 베이스는 여러 위치에서의 스피커 위치들의 조합, 예를 들어 그림 3 에서 (θ_1,θ_2)가 (-40, 10), (-30, 30) 또는 (-45, 45) 등 여러 경우에 대해서 적응 필터방법으로 구한다.

다. 역지연 함수의 적응 구현 방법

$T(z)$의 역지연행렬 $C(z)$는 기존의 Multiple Filtered-X LMS 방법을 사용하여 적응적으로 구할 수 있다. 그러나 본 논문에서는 계산의 효율성을 높이기 위해 다음과 같은 조금 다른 방법을 시도하였다.

식 (2)로부터 $T(z)$의 역행렬을 수식적으로 표현하여 식 (5)에 대입하면 $C(z)$는

$$C(z)=\frac{z^{-d}}{H_L(\theta_1,z)H_R(\theta_2,z)-H_L(\theta_2,z)H_R(\theta_1,z)}X\begin{bmatrix}-H_L(\theta_1,z) & H_R(\theta_1,z)\\ H_L(\theta_2,z) & -H_R(\theta_2,z)\end{bmatrix} \tag{7}$$

여기에서 $T(z)$ 행렬식의 역지연에 해당하는 $Q_0(z)$를 식 (8)과 같이 정의하면 $C(z)$는 식 (9)와 같이 된다.

$$Q_0(z)=\frac{z^{-d}}{H_L(\theta_1,z)H_R(\theta_2,z)-H_L(\theta_2,z)H_R(\theta_1,z)} \tag{8}$$

$$C(z) = Q_0(z)\begin{bmatrix} -H_L(\theta_1,z) & H_R(\theta_1,z) \\ H_L(\theta_2,z) & -H_R(\theta_2,z) \end{bmatrix} \quad (9)$$

$Q_0(z)$는 단일입력 단일출력 시스템이므로 Multiple Filtered-X LMS 방법대신 Filtered-X LMS 방법을 사용할 수 있다. 그림 5는 $Q_0(z)$를 Filtered-X LMS 방법을 적용하여 적응시키는 블록선도를 보여주고 있다. (θ_1, θ_2)에 대한 2행2열의 $C(z)$ 데이터베이스는(θ_1, θ_2)에 대한 $Q_0(z)$로부터 식 (9)를 이용하여 구축할 수 있다.

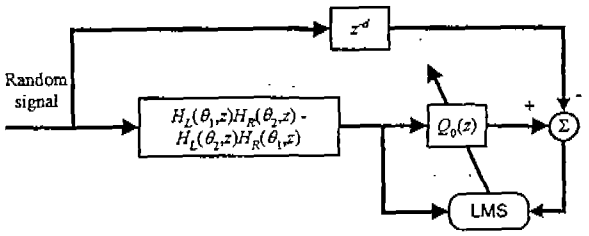

그림 5. $Q_0(z)$의 적응구조

라. 크로스톡 제거기 구성

앞서 구축된 고정 반경 경우의 $C(z)$의 데이터 베이스를 이용하면, 일반 위치의 스피커에 대한 크로스톡 제거기 $Q(z)$는 식 (5)를 이용하여 쉽게 구해진다. 그림 6은 그림 2의 크로스톡 제거 구성기를 앞서 기술한 흐름에 의해 자세히 도식화한 그림이다. 이 부분은 크게 두 부분으로 나누어져있다. 하나는 역지연 행렬 $C(z)$를 데이터 베이스화 하는 데이터베이스 구성부이며, 다른 하나는 이러한 데이터베이스와 유저인터페이스를 사용하여 실시간으로 크로스톡 제거기를 생성하는 부분이다.

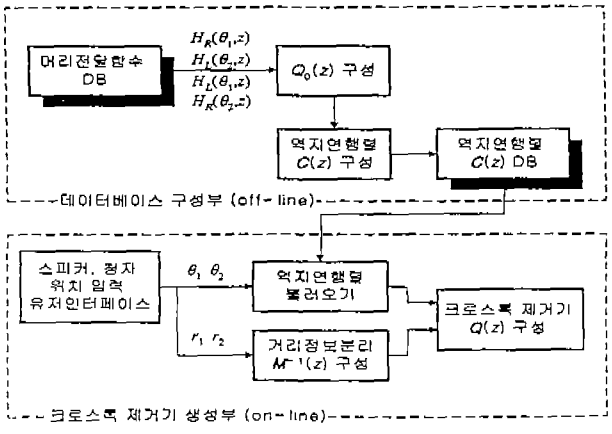

그림 6. 크로스톡 제거 구성기

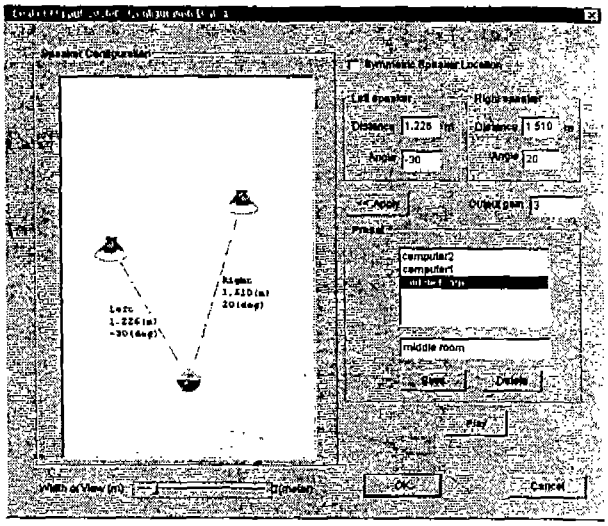

그림 7. 크로스톡 제거기 유저 인터페이스

유저 인터페이스

사용자로부터 2개의 스피커위치와 사용자의 청취위치를 받아들이는 유저인터페이스 부분이 그림 7에 있다. 참고로 이 유저인터페이스는 정보통신부 과제 '멀티미디어 콘텐트용 입체음향 처리 S/W 개발'의 일부이다.

몇 가지 자주 쓰이는 스피커-청취자 위치는 등록하여 재 사용할 수 있도록 되어있다. 또한 재생하고자 하는 스테레오 음향 파일을 유저인터페이스에 넘기면 이를 처리하여 스피커 재생용 신호로 변환해주는 기능도 가지고 있다.

결론

2스피커 재생시스템에서 스피커가 고정반경에 위치해 있는 경우에 대한 부분과 거리에 관련된 부분의 곱으로 분리하여 스피커-청취자 모델을 표현하였다. 제시된 방법은 스피커가 고정반경에 위치한 경우에 대한 역지연행렬만을 데이터베이스화 함으로써 일반위치 스피커의 경우까지 확장할 수 있어 매우 효율적으로 크로스톡 제거기 구성할 수 있다. 또한 데이터베이스 구축 시 2행2열의 직접적인 역지연행렬을 구하지 않고 단일 함수의 역지연을 구하였기 때문에 Multiple Filtered-X LMS 방법 대신 Filtered-X LMS 방법을 사용하여 적응 시 필요한 계산량을 약 75% 감소할 수 있었다. 이러한 일련의 과정을 쉽게 이용할 수 있는 유저인

터페이스를 개발하였으며, 여기에서는 스피커의 위치와 청취자의 위치를 입력함으로써 크로스톡 제거기가 즉각 구성되도록 하였다.

참고문헌

[1] 이희종, 서상원, 이명진, 김용완, "입체음향 기술", 소프트웨어 기술동향, 제 2 권 제 2 호, 시스템공학연구소, pp 54-80, 1998 년 6 월

[2] Ole Kirkeby and P. A. Nelson, "Properties of Least Squares Inverse Filters Used for Multi-channel Sound Reproduction", ACTIVE95, pp 1259-1270, 1995

[3] P. A. Nelson, H. Hamada and S. J. Elliott, "Inverse Filters for Multi-channel Sound Reproduction", IEEE Trans SP Vol. 40, No.7 pp 1621-1632, July, 1992

[4] S. J. Elliott, I. M. Stothers, and P. A. Nelson, " A Multiple Error LMS Algorithm and Its Application to the Active Control of Sound and Vibration," IEEE, Trans ASSP. Vol. 35, No.1, pp1423-1434, 1987