한국음향학회지 제29권 제8호 pp. 509~515 (2010)

비음수 행렬 분해와 디코릴레이터를 이용한 모노-스테레오 블라인드 업믹스 기법

Mono-To-Stereo Blind Upmix Using Non-Negative Matrix Factorization and Decorrelator

최 근 우*, 전 상 배*, 이 석 진*, 성 굉 모*
(Keunwoo Choi*, Sang Bae Chon*, Seokjin Lee*, Koeng-Mo Sung*)

*서울대학교 전기컴퓨터공학부 음향공학연구실
(접수일자: 2010년 9월 6일; 수정일자: 2010년 10월 12일; 채택일자: 2010년 11월 15일)

본 논문은 충분한 음원 너비 (Apparent Source Width)와 스테레오 이미지 품질 (Stereophonic Image Quality)을 확보하는 모노-스테레오 업믹스 기법을 제안한다. 모노 신호의 분석을 위해 높은 계수의 비음수 행렬 분해가 사용된다. 그 결과로 나온 분해된 음원들은 음조성 (Tonality)에 의하여 타악기 (Percussive)와 음조 (Tonal) 그룹으로 분류된다. 두 그룹 중 하나는 바로 스테레오 채널로 들어가는 반면 나머지 하나는 디코릴레이터를 통과하여 들어가게 된다. 청취 평가 결과 제안한 방법은 충분한 음원 너비와 스테레오 음상을 제공할 뿐만 아니라 기존의 방법에 비해 음색 변화도 감소하는 종합적으로 향상된 성능을 보여주었다.
핵심용어: 비음수행렬분해, 업믹스
투고분야: 음향 신호처리 분야 (1)

This paper presents a new method for upmixing mono signal to stereo signal with guaranteeing high stereophonic image quality (SIQ) and large apparent source width (ASW). The proposed method consists of analysis phase and synthesis phase. In analysis phase, a mono signal is first decomposed into multiple sound sources by the use of high-rank nonnegative matrix factorization. Then the multiple sources are clustered into two groups based on tonality criterion. In synthesis phase, one group is directly fed into left and right channels while the other group is decorrelated before being fed into each channel. Subjective tests reveals that the proposed method gives listener high SIQ and large ASW with minimizing timbral distortions.
Keywords: NMF, Mono Upmix
ASK subject classification: Acoustic Signal Processing (1)

I. 서 론

다양한 멀티채널 오디오 포맷의 등장으로 인하여 음원의 포맷 변환, 소위 '업믹스' 및 '다운믹스'의 필요성은 계속 증가하고 있다. 채널 수가 더 많은 포맷에서 적은 포맷으로의 변환인 다운믹스는 많은 정보를 상쇄 간섭 없이 잘 살리는 것이 중요한 화두이다. 반면 채널 수를 늘리는 업믹스에서 중요한 것은 주로 음상 정위를 통한 공간감의 구현이다. 채널 수가 늘어날수록 청자에게 더 다양하고 정확한 음상을 제공할 수 있는데, 이를 어떻게 활용할 것인지가 업믹스 알고리즘의 핵심 요소라고 할 수 있다. 한편, 업/다운믹스 알고리즘 중 '블라인드'는 이러한 과정이 음상이나 소스에 대한 사전 지식이 전혀 없는 상태에서 수행된다는 의미이다. 이런 점을 볼 때, 모노 블라인드 업믹스는 업믹스 중에서도 가장 극단적인 문제에 속한다. 모노 신호에는 기본적으로 여러 채널을 이용한 음상 정보가 존재할 수 없기 때문이다.

따라서 모노-스테레오 블라인드 업믹스 기법은 제시된 바가 거의 없다. Schroeder는 콤필터 디코릴레이터를 이용하여 유사 스테레오 효과를 내는 방법을 제시하였으며 [1], 다른 디코릴레이션 기법 역시 유사 스테레오 효과를 위하여 사용될 수 있다. 하지만 이 방법은 어떤 디코릴

책임저자: 최 근 우 (gnu@acoustics.snu.ac.kr)
151-744 서울시 관악구 신림동 서울대학교 132동 302호
(전화: 02-880-8427)

레이션 기법을 이용하더라도 주파수 왜곡이 불가피할 뿐만 아니라 음상 정위가 이루어지지 않아 스테레오 믹싱을 통한 여러 효과 중 포위감 (Listener Envelopement)의 형성에만 그 효과가 치중되어있다. 최근에 제안된 비음수 행렬 분해 [2]를 통한 분석은 모노 신호를 다루는 효과적인 방법 중 하나로 주목받고 있으며 이를 이용한 모노-5.1 서라운드 업믹스 기법이 제안된 바 있다. Uhle는 비음수 행렬 분해를 이용하여 잔향 성분을 추출하는 알고리즘을 제안하였으나 전면 스피커 신호는 단순한 디코릴레이션을 이용하였다 [3]. 블라인드가 아닌 방법으로는 M. Lagrange의 Normalized cuts를 이용한 전면 스피커의 업믹스 방법이 제안된 바 있다 [4]. 이 방법은 화상신호처리에 사용되는 Normalized cuts을 이용하여 주파수 도메인에서 음원을 나누는 방식인데, 분리된 음원을 직접 원하는 위치에 배치하는 방식으로 매우 적극적인 참여를 필요로 한다.

본 논문에서는 비음수 행렬 분해를 통한 모노-스테레오 업믹스 알고리즘을 다룬다. 오디오 신호 처리에서 비음수 행렬 분해는 주로 자동 악보 전사에 많이 쓰이는데 이는 신호를 주로 음표 별로 분류하는 비음수 행렬 분해의 특징 때문이다. 그런데 주어진 구간에 존재하는 음표보다 많은 계수 (Rank)를 이용하여 비음수 행렬 분해를 수행할 경우, 분해된 성분은 타악기 그룹과 음조 그룹으로 나눌 수 있다. 이러한 알고리즘은 이미 악보 전사에 사용된 바 있으며 [5], 본 논문에서는 타악기를 판별하는 방법으로 조성성을 이용하였다. 이 두 그룹 중 하나는 바로 스테레오 채널에 넣는데, 그 결과 해당 그룹은 청취자 정면에 음상이 맺히게 된다. 반면 또 다른 그룹은 디코릴레이터 필터를 통과시켜 포위감을 준다. 그 결과 전체적으로는 충분한 음원 너비와 음상 정위를 제공하게 되며 동시에 디코릴레이터의 사용으로 인한 음색 왜곡을 줄일 수 있다.

본 논문의 구성은 다음과 같다. II장에서는 비음수 행렬 분해와 조성도 계산, 디코릴레이터에 대한 간단한 설명을 제시한다. III장에서는 본 논문에서 제안하는 알고리즘에 대해 기술하고, IV장에서는 청취 평가와 그 결과 및 분석을 하였다.

II. 배경 이론

본 장에서는 제안한 알고리즘에 사용되는 비음수 행렬 분해와 디코릴레이터 및 음조성 계산에 대해 설명한다.

2.1. 비음수 행렬분해

비음수 행렬 분해는 Lee와 Seung에 의해 처음으로 소개되었다 [2]. 비음수 행렬 분해는 $n \times m$ 행렬 **V**를 $n \times r$ 행렬 **W**와 $r \times m$ 행렬 **H**의 곱으로 근사적으로 표현하는 것이 그 목표이고, 여기서 r은 계수 (Rank)가 된다. 즉, 수식 (1)을 만족하는 행렬 **W**와 **H**를 구하는 것이다.

$$V \approx WH \tag{1}$$

여기서 **W**와 **H**는 **V**와 **WH**간의 오차를 비용 함수를 통해 구하여 이를 줄여나가는 방법으로 반복적으로 계산된다. [2]에서는 Euclidean (EUC) distance와 Kullback-Leibler Divergence (KLD) 두 가지 비용함수가 반드시 수렴함을 증명하였다. 이 중 비교적 빠른 수렴 속도를 가지는 KLD가 널리 쓰이며 [6], 본 논문에서도 비용함수로 KLD를 이용하였다. KLD는 행렬 **A**, **B**의 오차를 수식 (2)와 같이 구한다.

$$D_{KL}(A,B) = \sum_{i,k} a_{ik} \log \frac{a_{ik}}{b_{ik}} - a_{ik} + b_{ik} \tag{2}$$

위의 식 (2)의 A가 **V**, B가 **W**와 **H**가 되며 이 비용함수를 줄이기 위한 적응 갱신 규칙은 다음과 같다 [2].

$$h^{+}_{ik} = h_{ik} + \alpha \left[\sum_{l=1}^{n} \frac{w_{li} v_{lk}}{(WH)_{lk}} - \sum_{l=1}^{n} w_{li} \right] \tag{3}$$

$$w^{+}_{ik} = w_{ik} + \alpha \left[\sum_{l=1}^{m} \frac{v_{il} h_{kl}}{(WH)_{il}} - \sum_{l=1}^{m} h_{kl} \right] \tag{4}$$

2.2. 조성도 (Tonality) 계산

단시간 푸리에 변환된 상태로 조성도를 구할 수 있는 방법이 제시되어 MPEG 부호화 기법 등 다양한 분야에서 응용되고 있다 [7]. 본 논문에서도 같은 방법으로 조성도를 구하였다. 신호 x [n]을 단시간 푸리에 변환한 것을 X (l,k) 라고 하고 그 크기와 위상을 각각 V (l,k)와 Φ (l,k)라고 하자. 여기서 l은 시간 지수 (index), k는 주파수 지수이다. 즉,

$$V(l,k) = |X(l,k)| \tag{5}$$

$$\Phi(l,k) = \angle X(l,k) \tag{6}$$

이다. 이 경우 과거 2탭을 이용하여 현재의 크기와 위상을 다음과 같이 예측할 수 있다.

$$\hat{V}(l,k) = V(l-1,k) + (V(l-1,k) - V(l-2,k)) \quad (7)$$

$$\hat{\Phi}(l,k) = \Phi(l-1,k) + (\Phi(l-1,k) - \Phi(l-2,k)) \quad (8)$$

이 값과 실제 값과의 비교를 통해 무질서도 (Chaos measure)를 측정할 수 있다. 이는 다음과 같다.

$$c(l,k) = \frac{(\hat{V}(l,k) - V(l,k))^2 + (\hat{\Phi}(l,k) - \Phi(l,k))^2}{V(l,k) + |\hat{V}(l,k)|} \quad (9)$$

이 무질서도는 완전한 잡음의 경우 최대 1, 평균 0.5의 값을 가지게 되며 조성도가 강한 경우에도 대부분의 0.05 이상의 값을 가진다는 특징을 이용해 다음과 같이 범위를 제한하고, 이를 로그 함수를 이용해 ()와 같이 표현하면 최종적으로 조성도를 구할 수 있다.

$$c_l(l,k) = \max\{0.05, \min[0.5, c(l,k)]\} \quad (10)$$

$$T(l,k) = -0.43 \times \log 10(cl(l,k)) - 0.299 \quad (11)$$

결과적으로 조성도는 최소 −0.17, 최대 2.6의값을 갖도록 정의되었다 [7].

마지막으로, 조성도의 계산은 기본적으로 단시간 푸리에 변환의 창함수 (window) 및 중복 구간의 길이에 영향을 받는다는 점을 고려할 필요가 있다

2.3. 디코릴레이터

본 논문에서는 Hawksford가 제안한 디코릴레이터를 이용하였다 [8]. 이 방법은 필요한 채널 수 만큼의 백색 잡음 신호를 만든 뒤 그 신호를 주파수상에서 등화 (Equalize)을 하여 각각의 입력 신호와 컨볼루션을 하는 방식이다. 이 방식은 작동이 빠를뿐만 아니라 충분한 디코릴레이션 성능을 제공한다. [8]에서는 다양한 창함수 모양을 제시하였는데 그 중에서 지수형 감쇠 창함수를 채택하였다. 지수형 감쇠 창함수는 감쇠 계수의 조절을 통해 디코릴레이션의 조절이 용이한 장점이 있다 [8].

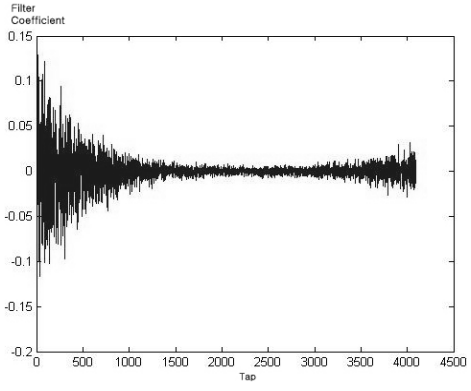

그림 1. 지수 감쇠형 창함수를 이용한 디코릴레이션 필터
Fig. 1. Decorrelation filter using exponentially decaying window.

III. 알고리즘 제안

3.1. 비음수 행렬 분해를 통한 성분분해

비음수 행렬 분해를 하려면 기본으로 비음수성을 가정한다. 이는 신호를 단시간 푸리에 변환 한 뒤 그 크기만을, 즉 스펙트로그램 매그니튜드 V를 다룸으로써 가능해진다. 본 연구에서는 46.4ms (2048 탭) 길이의 해밍 (Hamming) 창함수와 7/8 중복 구간을 적용하였고 푸리에 변환은 4096 포인트로 수행하였다. 일반적으로 비음수 행렬 변환을 통한 응용에서는 유사한 수준의 주파수 및 시간 해상도가 요구된다 [4, 9, 10].

분해된 성분은 식 (12)와 같이 구할 수 있으며, 분해하기 전의 스펙트로그램 매그니튜드 V와 같이 크기만을 가지고 있게 된다. 따라서 이를 오디오 신호로 복원하기 위한 위상 정보는 분해되기 전 신호의 위상을 이용한다. 즉, 분해된 성분 Vp는 두 행렬의 곱

$$V_p = W_p \times H_p \quad (12)$$

로 표현되며, 위상 정보까지 포함한 최종 분해 성분은

$$X_p(l,k) = V_p(l,k) \times \angle X(l,k) \quad (13)$$

라고 표현된다. 식 (13), (13)에서 p는 1부터 계수 R 까지의 정수로, 총 R개의 성분을 구할 수 있다.

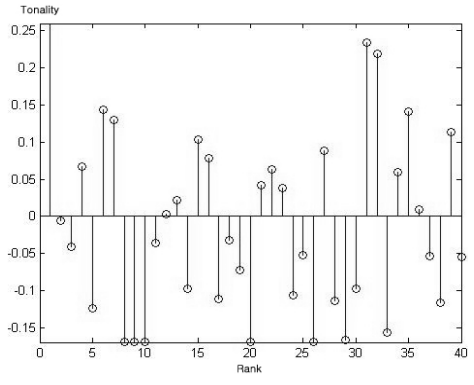

그림 2. Michael Jackson의 'Billie Jean'의 성분 40개에 대한 조성도
Fig. 2. The tonalities of 40 components from 'Billie Jean', Michael Jackson.

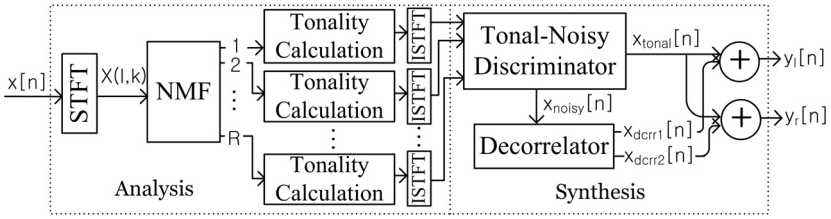

그림 3. 제안한 업믹스 알고리즘 중 알고리즘 A의 블록선도
Fig. 3. Block diagram of proposed upmix algorithm A.

3.2. 조성도 계산

2.2에서 설명한 바와 같이, 식 (5)에서부터 (11)까지의 과정을 거쳐 조성도를 계산하였다. 이 계산 방법에 의해 조성도를 계산할 경우, 조성도는 무음 구간에서 최소값을 갖게 되므로 무음 구간이 없다고 가정한다. 하지만 비음수 행렬 변환을 통해 나온 성분들은 대부분 많은 무음 구간을 가지게 되므로 본 논문에서는 음이 지속되는 구간 중 에너지가 가장 큰 구간을 구하여 그 부분에 대해서만 조성도를 구하였다.

타악기 및 음조 그룹을 나누는 기준은 −0.1부터 0.1 사이의 다양한 값을 넣어 분리한 뒤 들었을 때의 조성도 및 자연스러운 정도를 통하여 최종적으로 0.025로 정하였다. 그 결과 실험한 6개의 음원에 대하여 총 40개의 성분 중 평균적으로 14.7개가 타악기 그룹에 속하였고 25.3개가 음조 그룹에 속하게 되었다.

3.3. 디코릴레이터의 적용

최종적으로 나뉜 그룹 중 하나만 디코릴레이션 필터를 통과하게 된다. 여기서는 2.3에서 설명한 디코릴레이터 필터를 사용하였다. 93 ms (4096탭)의 백색 잡음 필터가 사용되었다. 길이가 길수록 디코릴레이션이 강하게 일어나는 반면 음색의 변화가 심해진다.

청취 평가에서 기존 기법과의 비교를 위해 만든 음원에도 똑같은 디코릴레이터를 적용하였으며 93 ms 의 길이는 적당한 수준의 모조 스테레오 효과를 내기에 적합한 길이이다.

전체 알고리즘의 블록선도를 그림 3에 표현하였다. 이는 알고리즘 A의 블록 선도로, 타악기 그룹에 디코릴레이터를 걸어준다. 알고리즘 B의 경우에는 이와 반대로, 음조 그룹이 디코릴레이터를 통과한다. 이에 대한 설명은 3.4에 언급되어있다.

3.4. 알고리즘 A와 B

제안한 알고리즘 중 알고리즘 A는 그림 3과 같이 타악기 성분을 디코릴레이션시킨 반면 음조 성분은 L, R 채널에 그대로 보내어 결과적으로 음조 성분의 음상을 중앙에 정위시킨 방법이다. 한편 알고리즘 B는 이와 반대로 타악기 성분의 음상을 중앙에 정위시키고 음조 성분을 디코릴레이션 시킨 것이다. 자연스러운 음상정위와 음색 변화에 대한 주파수 민감성 등의 효과를 고려할 때 두 가지 방법의 우열은 청취평가를 통해 밝혀내기 위하여 두 가지 방법을 모두 평가하였다.

IV. 청취 평가 및 결과

4.1. 평가 항목의 결정

평가 항목 중 공간감에 해당하는 요소로는 정확하고 안정적인 음상 정위를 평가하기 위한 스테레오 이미지 품질 (Stereophonic Image Quality, SIQ) [11]과 포위감에 직접적인 영향을 주는 음원 너비 (Apparent Source Width, ASW) [12, 13]를 선정하였고 이와 별도로 음색을 평가하였다. 이렇게 나누어 평가하는 이유는 음색과 공간감은 평가에 있어서 서로 독립적인 요소이기 때문이며 [14], 공간감에 있어서도 ASW와 SIQ는 상반된 성격을 띄고 있기 때문이다. 마지막으로 이 세 가지 평가 항목을 이용하여 종합 점수 (Basic Audio Quality, BAQ) [11, 14]를 산출하였다.

4.2. 평가 방법

전체 21명의 청취자가 실험에 참여하였고, 락, 팝 및 재즈를 포함한 6개의 곡을 선정하여 그 중 8초 분량을 비교하였다. 모든 음원은 44,100의 표본율과 16비트 양자화로 복호화된 PCM WAV 파일이다. 실험은 0.3초의 잔향 시간을 갖는 시청실에서 이루어졌으며 음원 재생을 위해 Genelec 8050A가 사용되었다. 청취자는 Sweet spot에 앉아 실험을 진행하였다.

우선 ASW와 SIQ를 평가하는 실험 1, 2 에서는 쌍비교 방식을 이용하였다. 여기에서는 스테레오 원본, 기존 방

식, 알고리즘 A, 알고리즘 B 및 모노 신호를 이용하였다. 한편 음색 평가를 위한 실험 3에서는 MUSHRA 방식을 이용하였고, 스테레오 원본, 모노신호, 알고리즘 A, 알고리즘 B, 기존 방식 및 로우 패스 필터를 통과시킨 앵커를 이용하였다.

4.2.1. 쌍비교 (Paired Comparison)을 이용한 공간감 평가

전문적인 음악 엔지니어와는 달리, 일반적인 청취자들은 음상 정위의 변화에 둔감한 반면 포위감을 선호하는 특징이 있다 [15]. 따라서 ASW와 SIQ를 분리하여 평가하였다. 청취자는 각 음원별로 5개 중 임의로 선택한 2개의 다른 신호를 비교하는데 여기에는 음악 엔지니어가 믹싱한 스테레오 음원 원본이 포함되어있다. 최종적으로는 ASW와 SIQ 평가에서 각각 $6\times {}_5C_2 = 60$개의 쌍비교를 수행하게 된다. 점수 스케일은 -3부터 3까지 1점 간격이다.

ASW와 SIQ는 어느 정도 Trade-off의 관계가 있으며, 음색 변화 또한 커지게 된다. 따라서 평가에 있어서 원본보다 ASW가 더 커질수는 있으나, 그것을 업믹스 알고리즘이 추구할 방향이라고 할 수는 없다. 그러므로 ASW는 원본에 가까운 수준이면 충분히 달성되었다고 판단할 수 있다. 반면 SIQ는 원본을 뛰어넘는 것은 현실적으로 불가능하므로 높을수록 좋다고 할 수 있다. 만일 이런 특징을 무시하고 모든 항목을 MUSHRA (MUlti Stimulus with Hidden Reference and Anchor) [16]를 통해 평가할 경우, ASW는 기존 방법과 같이 단순히 디코릴레이션을 강하게 걸기만 하면 높은 점수를 받는 반면 스테레오 원본이나 다른 방법은 적당한 수준의 ASW를 가지고 있더라도 기존 방법과 비교되어 다같이 낮은 점수를 받게 되므로 변별력이 떨어진다. SIQ는 반대 경우인데, 기존 방법은 SIQ가 매우 떨어지므로 상대적으로 다른 방법간의 구별이 모호해진다. 따라서 이 두 경우에는 쌍비교를 통해 평가하는 것이 적합하다.

일반적으로 사용되는 가장 단순한 쌍비교 결과 처리는 점수를 모두 더하는 것인데 [17], 이 방법은 N회 비교하여 모두 이기는 경우 1.5×N점을, 모두 지는 경우 -1.5×N점을 부여한다. 여기서는 모든 점수를 더한 뒤 전체에 1.5*N점을 더하여 최종적으로 모두 이긴 경우 3N점을, 모두 진 경우 0점을 받게 가공한 뒤 스테레오 원본이 100점이 되도록 정규화하였다.

4.2.2. MUSHRA를 이용한 음색 평가

음색 평가에 있어서는 위와 같은 문제가 일어나지 않는다. 즉 스테레오 원본을 참조 신호로 하여 평가하면 정확한 음색 평과 결과를 얻을 수 있으므로 MUSHRA를 이용하였다. 이는 효율적면서도 정확한 결과를 얻을 수 있게 해준다.

4.2.3. 공간감과 음색 점수를 통한 종합점수 (Basic Audio Quality)

식 (14)는 [14]에서 제안한, BAQ에 기여하는 음색과 공간감의 비율을 계산하여 종합점수를 구하는 식이다 [14]. 에서는 공간감 점수를 전면과 서라운드로 나누었으나 본 논문은 스테레오 신호의 평가로 서라운드 점수가 존재하지 않는다. 대신 공간감을 ASW와 SIQ로 나누어 평가하였으므로 이를 고려하여 공간감 점수의 계산을 식 (15)와 같이 수정하였다. 음색 점수에 해당하는 Timbral은 4.2.2에서 설명한 방법을 이용하여 얻은 점수를 사용한다.

$$BAQ = 0.80Timbral + 0.39Spatial - 18.7 \tag{14}$$

$$Spatial = \sqrt{ASW_n^2 + SIQ^2} \tag{15}$$

$$ASW_n = \min(ASW, 100) \tag{16}$$

이는 ASW와 SIQ 두 요소에 같은 가중치를 부여하여 평가한 것이고, 두 요소에서 고르게 점수를 받는 것이 하나만 잘 받는 것 보다 좋은 공간감을 준다고 가정하여 (ASW, SIQ)의 크기를 전체 공간감 점수로 사용하였다. 마지막으로 식 (16)은, ASW의 경우 100을 넘는, 즉 스테레오 원본보다 높은 ASW를 받는 것은 의미가 없다는 것을 고려한 식이다. 음색 점수는 MUSHRA 결과를 사용하였다.

4.3. 평가 결과

그림4은 ASW와 SIQ의 평가 결과이다. 예상한 바와 같이 기존 방법, 즉 단순히 디코릴레이션만을 이용한 경우에 ASW는 매우 높은 점수를 받았는데, 4.2.1에서 설명한 바와 같이, 스테레오 원본이 100점을 받도록 정규화한 결과로 100점이 넘는 점수는 스테레오 원본보다 다소 높은 ASW를 가진 것을 의미한다. 한편 제안한 방법은 스테레오 원본에 비하여 72.5와 85.1로 ASW 측면에서는 다소 모자란다고 할 수 있다. 하지만 기존 방법은 SIQ에서 26.5를 받은 반면 제안한 방법은 각각 39.9와 30.7라는

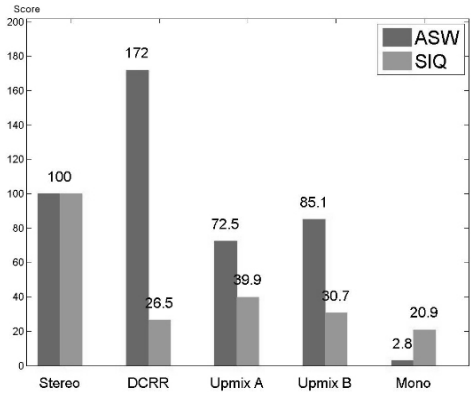

그림 4. ASW 및 SIQ 점수
Fig. 4. Average scores for ASW and SIQ.

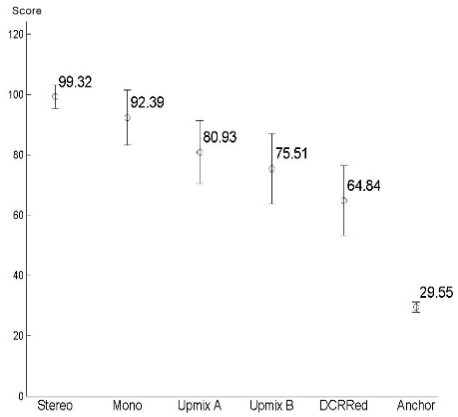

그림 5. 음색 평가 결과 (평균 및 95 % 신뢰구간)
Fig. 5. Average scores and 95 % confidence intervals for timbre.

향상된 평가를 받았다. 모노 신호가 20.9를 받은 것을 고려하면 기존 방법에 비해 많은 향상을 보인 셈이다.

그림 5는 음색 평가 결과이다. 스테레오 원본을 참조 신호 (Reference)로 하여 평가한 이 결과에서 제안한 방법은 각각 80.93과 75.51이라는 점수를 받아 64.84를 받은 기존 방법에 비해 향상된 성능을 보인다. 한편 이를 식 (14)을 이용하여 평가한 종합점수는 다음과 같다. 이는 타악기 혹은 음조 그룹에만 디코릴레이터를 적용한 결과가 음원 전체에 디코릴레이터를 적용한 것 보다 더 원본에 가까운 음색을 가지고 있음을 보여준다. 특히 알고리즘 A가 더 높은 점수를 받았는데 이는 음조 그룹의 음색에 대해 청취자가 더 민감하게 반응하기 때문으로 보인다.

정리하면, 제안한 알고리즘 A와 B는 ASW에서 기존 방법보다 낮지만 스테레오 원본에 근접한 점수를 받았다. 반면 SIQ에 대해서는 기존 방법보다 높은 점수를 받았으며 음색 왜곡도 덜 발생하는 것을 확인할 수 있었다. 종합점수에서는, 스테레오 원본에 이어 알고리즘 A가 가장

표 1. 각 방법 별 종합점수 결과
Table 1. Basic audio quality scores comparison.

Stereo	Conventional	Upmix A	Upmix B	Mono
99.8	61.7	68.9	66.7	61.0

높은 점수를 받았으며, 알고리즘 B도 기존 방식보다 높은 점수를 받았다. 기존의 방법은 모노 신호에 비해 큰 차이가 없는 점수를 받았다. 이는 공간감의 향상이 기여하는 만큼이나 음질 열화가 크다는 것을 의미한다.

음조 그룹의 음상을 중앙에 정위시키고 타악기를 디코릴레이션 필터에 통과시킨 알고리즘 A가 더 좋은 평가를 받은 이유는 여러가지로 분석할 수 있다. 우선 알고리즘 A는 음색 측면에서 알고리즘 B보다 높은 점수를 받았다. 디코릴레이션 필터를 통과하며 일어난 음색 변화를 고려하였을 때, 음조 그룹에 포함된 보컬이나 기타, 피아노 등의 음색 변화에 청취자들이 더 민감하게 반응했다고 볼 수 있다. SIQ도 알고리즘 A가 알고리즘 B보다 높은 점수를 받았다. 청취자들은 음상 정위가 모호한 경우 메인 멜로디가 자신의 정면에 있다고 가정하는 경향을 보였는데, 그 결과 알고리즘 B에서는 중앙에 음상을 정위한 타악기 그룹뿐만 아니라 디코릴레이션 필터를 통과한 멜로디 역시 중앙에 있다고 느끼게 되어 SIQ가 낮게 나왔다고 해석된다. 반면 알고리즘 A에서는 처음부터 음조 그룹의 음상이 중앙에 정위되어 있어 이런 현상이 일어나지 않아 결과적으로 더 높은 SIQ를 받았다.

V. 결 론

본 논문에서는 새로운 방식의 모노-스테레오 블라인드 업믹스 알고리즘을 제안하였다. 제안한 알고리즘은 ASW, SIQ, 그리고 음색 세가지 면에서 종합적으로 좋은 성능을 발휘하기 위해 설계되었다. 이를 위해 비음수 행렬 분해를 통한 성분 분석과 조성성 기준에 의한 부분적인 디코릴레이션이 이루어졌다. 또한, 쌍비교 방법과 MUSHRA를 이용하여 청취 평가를 시행하였으며, 해당 청취 평가를 통해 제안한 알고리즘이 기존의 업믹스 알고리즘과 비교하여 전체적으로 향상된 성능을 가짐을 확인할 수 있었다. 특히 타악기 그룹을 디코릴레이션시키고 음조 그룹을 가운데 위치시킨 방법이 가장 높은 점수를 받은 것은 주목할 만한 결과이다..

본 논문에서 제안한 알고리즘은 모노 신호로 녹음된 음악의 재생에 활용될 수 있을 것으로 기대된다.

참 고 문 헌

1. M. R. Schroeder and B. F. Logan, "Colorless artificial reverberation," *J. of AES*, no. 3, pp. 192–197, 1961.
2. D. Lee and H. Seung, "Algorithms for non–negative matrix factorization," in *Proc. NIPS*, 2001.
3. C. Uhle, A. Walther, and M. Ivertowski, "Blind one–to–N upmixing," *AudioMostly 2nd Conference*, pp. 110–115, September, 2007.
4. M. Lagrange, L. G. Martins, and G. Tzanetakis, "Semi–automatic mono to stereo up–mixing using sound source formation", *AES 125th Convention*, paper no. 7042, May, 2007.
5. M. Helen, "Separation of drums from polyphonic music using non–negative matrix factorization and support vector machine," *EURASIP 13th Conference*, September, 2005.
6. E. Benetos, "Musical instrument classification using non–negative matrix factorization algorithms and subset feature selection," in *Proc. IEEE Conference on Acoustics, Speech, and Signal Processing*, May, 2006.
7. K. Brandenburg and J. D. Johnston, "Second generation perceptual audio coding : The hybrid coder," *AES 88th Convention*, March, 1990.
8. M. O. J. Hawksford and N. Harris, "Diffuse signal processing and acoustic source characterization for applications in synthetic loudspeaker arrays," *AES 122nd Convention*, April. 2002.
9. C. Uhle, "Ambience separation from mono recordings using non–negative matrix factorization," *AES 30th Conference*, March, 2007.
10. P. Smaragdis, "Non–negative matrix factorization for polyphonic music transcription," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October, 2003.
11. ITU–R (1997). *Recommendation BS. 1116–1 : Recommendation BS. 1116: Methods for subjective assessment of small impairments in audio systems including multichannel sound systems*, International communication union.
12. M. Morimoto, "The role of rear loudspeaker in spatial impression", *AES 103th Convention*, paper no. 4554, September, 1997.
13. F. Rumsey, "Subject assessment of the spatial attributes of reproduced sound," *AES 15th Conference, October*, 1998.
14. F. Rumsey, S. ZielińČski, and R. Kassier, "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," *J. of ASA*, vol. 118, Issue 2, pp. 968–976, August, 2005.
15. F. Rumsey, "Spatial audio and sensory evaluation techniques – context, history and aims," *Spatial audio and sensory evaluation techniques conference*, April, 2006.
16. ITU–R (2001). *Recommendation BS. 1534–1 : Method for the subjective assessment of intermediate quality level of coding systems*, International communication union.
17. ITU–T (1996). *Recommendation P.800 : Method for object and subject assessment of quality*, International communication union.

저자 약력

•최 근 우 (Keunwoo Choi)

2009년: 서울대학교 전기공학부 졸업 (학사)
2009년 ~ 현재: 서울대학교 전기컴퓨터공학부 음향공학연구실 석사과정
※ 관심분야: 음향 신호처리

•전 상 배 (Sang Bae Chon)

2000년: 전북대학교 전기전자 제어공학부 학사졸업
2005년: 서울대학교 전기컴퓨터 공학부 석사졸업
2009년: 서울대학교 전기컴퓨터 공학부 박사졸업
※ 관심분야: 음향학, 음향 및 음성 신호처리

•이 석 진 (SeokJin Lee)

2006년 8월: 서울대학교 공과대학 전기·컴퓨터공학부 (공학사)
2008년 8월: 서울대학교 공과대학 전기·컴퓨터공학부 (공학석사)
2008년 9월 ~ 현재: 서울대학교 공과대학 전기·컴퓨터공학부 박사과정
※ 관심분야: 음향 신호처리

•성 굉 모 (Koeng–mo Sung)

1973년: 독일 아헨대학교 전기공학 학사졸업
1977년: 독일 아헨대학교 전기통신공학 석사졸업
1982년: 독일 아헨대학교 음향공학 박사졸업
1983년 ~ 현재: 서울대학교 전기컴퓨터공학부 정교수
※ 관심분야: 음향학, 음향 및 음성 신호처리