차원 축소를 이용한 주파수 영역 오디오 신호 압축

*김민제 백승권 이태진 장대영 강경옥
한국전자통신연구원
*mkim@etri.re.kr

Dimensionality Reduction Based Frequency Domain Audio Signal Compression Method

*Kim, Minje Beack, Seungkwon Lee, Taejin Jang, Daeyoung Kang, Kyeongok
Electronics and Telecommunications Research Institute

요약

본 논문은 오디오 부호화 및 복호화 과정에서, 주파수 영역에서 표현된 오디오 신호를 차원 축소 방법으로 압축하여 표현함으로서 오디오 부호화 효율을 증대시키고자 하는 방식에 관한 것이다. 차원 축소는 행렬을 특정한 조건을 바탕으로 두 개의 행렬의 곱으로 표현하는 방식으로, 특정 행렬로 표현된 데이터를 좀 더 작은 데이터량으로 표현하는 것뿐만 아니라 이 과정에서 데이터에 내재되어 있는 추상적인 정보까지도 함축적으로 얻어낼 수 있기 때문에, 일반적으로 데이터의 압축에 좋은 성능을 보인다. 주파수 영역으로 변환된 신호는 일반적으로 (주파수 밴드의 개수) × (전체 프레임의 개수)인 행렬로 볼 수 있으며, 이 전체 행렬을 입력으로 간주하고, 차원 축소를 수행하여 신호의 압축 효과를 얻을 수 있다. 그러나 이 경우, 행렬 전체를 입력 신호로 보아야 하기 때문에 실시간 부호화가 불가능하며, 신호 전체 길이만큼의 부호화 지연이 발생한다. 이를 해소하기 위해, 본 논문에서는 특정 개수만큼의 프레임을 묶어서 여러 번의 차원 축소를 순차적으로 수행함으로써 부호화 지연을 최소화하는 방식을 제안한다.

1. 서론

CD(Compact Disk) 혁명 이후 가속화된 오디오 신호의 디지털 표현은 “MP3”라는 확장자로 더 잘 알려진 MPEG 표준 규격, MPEG-1 Layer Ⅲ [1]로 인해 오디오 신호를 물리적 스토리지 형태에 구애 받지 않고 저장/전송하는 기술적 혁신을 이루었다. 오디오 신호의 디지털 표현 방식은 현재, MPEG-2 AAC(Advanced Audio Coding) [2]를 거쳐 낮은 비트율에서 음성 신호의 부호화시 낮은 성능을 보이는 오디오 신호 부호화기와 높은 비트율에서 오디오 신호의 부호화시 낮은 성능을 보이는 음성 신호 부호화기의 단점을 보완할 수 있는 통합 부호화기의 표준화가 진행되고 있다.

오디오 신호를 부호화하여 디지털 신호로 표현하는 것의 가장 큰 목적은 아날로그 형태의 원 신호와 같은 음질을 유지하면서도 가능한 한 적은 정보량으로 그 신호를 표현하는 것이다. 이는 신호를 방송, 통신망을 통해 전송하거나 저장함에 있어서 매체를 잠식하는 데이터의 양을 줄임으로써 보다 많은 정보를 취급하는 것을 목표로 하는 것이며, 이러한 신호 압축의 필요성은 방송 환경이 이동하는 사용자를 대상으로 확대되고, 일반 소비자의 음향 청취 환경이 다채널화되면서 더욱 증대되고 있다.

오디오 부호화 기술은 크게 신호가 가지고 있는 통계적 특성을 이용하여 원 신호를 열화 없이 압축 부호화하는 방식과, 인간의 청각적 특성을 이용하여 열화된 신호가 원 신호와 지각적으로 차이가 없도록 부호화하는 두 가지 방식으로 나눌 수 있다. 본 논문에서는 신호 자체가 가지고 있는 중복성을 차원 축소 알고리즘을 통해 드러나도록 유도함으로써, 주파수 영역으로 변환된 신호를 획기적으로 압축하여 표현할 수 있는 방식을 제안한다. 일정한 개수 이상의 데이터 샘플이 모여야만, 이를 분석하여 그 특질을 통해 데이터의 차원을 축소할 수 있는 차원 축소 방식들은 필연적으로 부호화/복호화시 지연을 발생시키게 되며, 본 논문은 이를 해소하기 위한 방식으로 스케일 윈도우를 통한 점진적 차원 축소 방식을 제안한다.

2. 오디오 신호의 주파수 영역 표현

현존 오디오 부호화 기술은 시간 영역 신호를 주파수 영역으로 변환하여 표현함으로써 부호화 이득을 취하고 있다. 현재의 오디오 부호화 기술에서 주로 사용되는 주파수 영역 변환 방식에는 크게 필터 뱅크를 이용하는 방식과 블록 변환 방식이 있다. [3]

가. 필터 뱅크

필터 뱅크를 이용한 주파수 영역으로의 변환은 그림 1 과 같이 도식화할 수 있다. 기본적인 방식은 시간 영역 신호를 필터 뱅크에 통과시킴으로써 K 개의 주파수 밴드로 분해하는 것이다. 이후 각각의 주파수 밴드 별 신호가 제한된 비트 수에 의해 양자화되며, 이 과정에서 대부분의 양자화 잡음은 심리음향모델을 통해 얻어진 마스킹 정보를 고려하여 가장 들리지 않는 주파수 밴드에 할당한다. 양자화된 신호는 복호화기로 보내지며, 이곳에서 주파수 밴드 별로 복호화된 신호는 전대역 신호를 복원하기 위해 합쳐진다. K개의 병렬적 밴드를 사용하는 이 방식에서의 가장 큰 문제점은, 밴드 별로 분할하면서 데이터 양이 K 배로 늘어난다는 점이고, 이 문제점을 해소하기 위해 밴드 별로 K 샘플마다 하나씩만의 샘플을 취하는 down sampling 방식을 적용해야 한다. 이 과정에서 손실되는 sampling rate 를 보전하기 위해 복호화 단

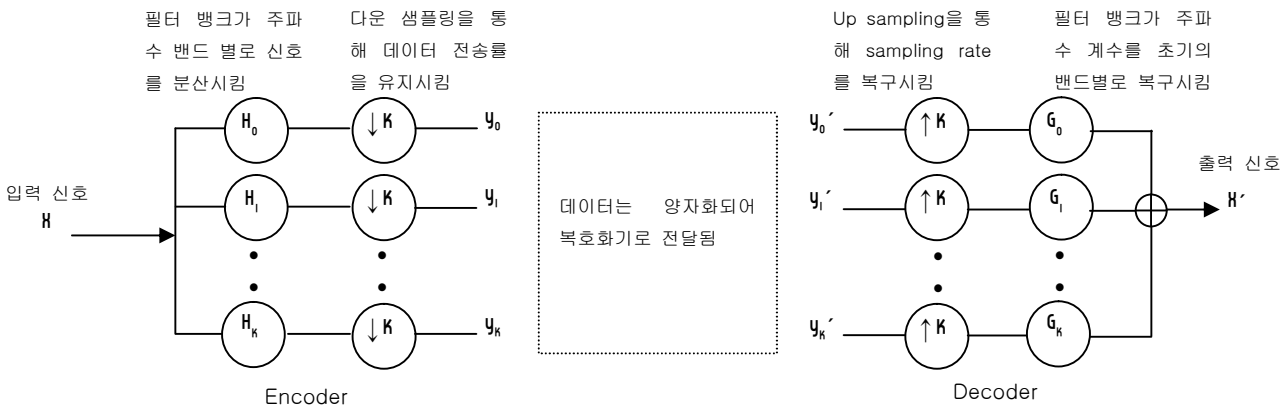

그림 1. 필터 뱅크를 이용한 시간-주파수 영역 변환

계에서 반대로 up sampling 을 수행해야 한다. PQMF(Pseudo Quadrature Mirror Filter) [4][5]의 경우 상용 오디오 부호화기(MPEG 1,2 Layers I, II, III)[6][1]에서 사용되는 필터 뱅크 방식의 시간-주파수 영역 매핑의 대표적인 예이다. PQMF 필터 뱅크는 K 개의 채널로 이루어져 있으며, 각각의 채널은 코사인에 의해 변조된 저대역 필터 $h[n]$이다. 분석 및 합성 필터의 형태는 식 1 과 같다.

$$h_k[n] = h[n]\left(\pi\left(\frac{k+\frac{1}{2}}{K}\right)\left(n-\frac{N-1}{2}\right)+\Phi_k\right) \quad \text{for } k=0,\ldots,K-1$$
$$g_k[n] = h_k[N-1-n] \qquad (1)$$

식 1 에서 N은 $h[n]$ 필터의 샘플 개수이며, 위상 정보 Φ_k는 인접 밴드 간 위신호를 막기 위해 식 2 와 같은 조건으로 결정된다.

$$\Phi_k - \Phi_{k-1} = \frac{\pi}{2}(2r+1) \qquad (2)$$

식 2 에서 r은 임의의 정수이다.

나. 블록 변환

시간-주파수 영역 매핑의 또 다른 주요한 방식은 블록 변환 방식이다. 블록 변환 방식과 필터 뱅크 방식은 서로 다른 발전 과정을 거쳐왔지만, 내부적인 기본 방식은 동일한 알고리즘이라고 볼 수 있으며,

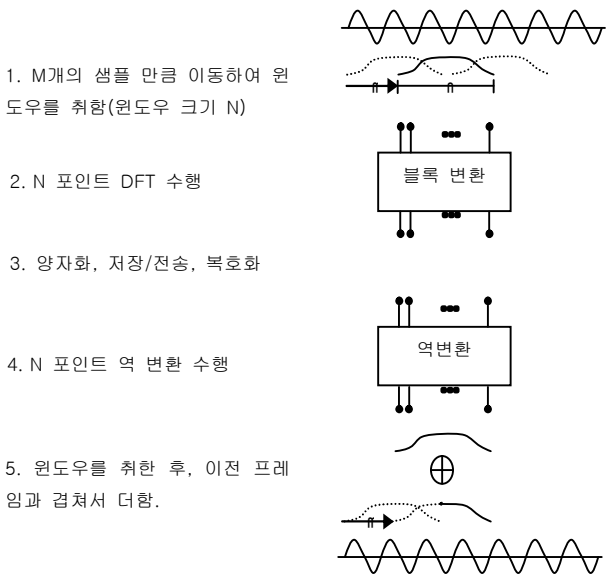

그림 2. 윈도우와 Overlap-and-add 방식을 이용한 오디오 데이터 부복호화 방식

둘의 가장 큰 차이는 블록 변환 방식이 더 많은 수의 밴드를 사용한다는 점이다. 이러한 차이점에 의해 많은 개수의 주파수 채널을 사용하는 오디오 부호화기(MPEG AAC [2], Dolby AC-2 와 AC-3 [7], AT&T/Lucent PAC [8] 등)는 블록 변환 방식을 사용한다.

이산 퓨리에 변환 (DFT: Discrete Fourier Transform)

밴드 별로 제한된 신호에 대해서는 sampling rate 가 최대 주파수의 2 배 이상이 된다면 이산적 신호 샘플로 표현할 수 있다. 반대로, 시간 도메인에서 제한이 있는 신호는 이산적인 주파수 샘플을 이용하여 완전히 표현할 수 있다. 시간/주파수 영역 모두에서 제한이 있는 신호는 시간/주파수 영역에서 모두 이산적으로 표현할 수 있다는 점을 이용하면, 퓨리에 변환을 이산 샘플로 표현할 수 있게 된다. 그림 2 는 윈도우 함수를 이용하여 시간 영역 신호를 제한시킨 다음 퓨리에 변환을 수행하여 제한된 이산 주파수 샘플을 만들고, 그 결과가 복호화기에서 역변환을 통해 다시 시간 영역 신호로 변환되는 과정을 나타낸다. 상기 과정에서 실제 전송되는 데이터는, 시간 영역에서의 정보가 아니라 주파수 밴드 별 복소수 계수가 전송되게 된다. 시간 영역 샘플을 주파수 영역으로 변환하는 방식과, 주파수 영역의 샘플을 시간 영역으로 역변환하는 방식은 식 3 과 같이 나타낼 수 있다.

$$X[k] \equiv \sum_{n=0}^{N-1} x[n]e^{-\frac{j2\pi kn}{N}} \qquad k=0,\ldots,N-1$$
$$x[n] \equiv \frac{1}{N}\sum_{n=0}^{N-1} X[k]e^{\frac{j2\pi kn}{N}} \qquad n=0,\ldots,N-1 \qquad (3)$$

수정된 이산 코사인 변환 (MDCT: Modified Discrete Cosine Transform)

블록 변환 부호화 과정에서는 블록 처리에 의한 잡음을 막기 위해 overlap-and-add 방식을 사용하는데, MDCT[9]는 이 때 필연적으로 발생되는 데이터량 증가 문제를 해소하면서도 여전히 블록 변환을 가능하게 해 주는 장점이 있다. 이는 시간 영역 데이터 샘플을 프레임 크기의 절반인 N/2 만큼씩 취하여 overlap-and-add 에서의 왼쪽 윈도우와 오른쪽 윈도우의 필터 역할을 동시에 수행해주는 방식이다. 그림 3 은 MDCT 방식에서 overlap-and-add 가 수행되는 과정을 나타내고 있다. 먼저, 각 N/2 개의 샘플을 포함하는 i 번째 블록 및 i-1 번째 블록은 하나의 블록으로 간주되어 각각 윈도우 함수의 오른쪽 부분 및 왼쪽 부분과 곱해진다. 이 결과는 커널 변환(A1 및 A2)을 통해 N/2 개의 주파수 샘플로 변환되며, 전송 또는 저장된 이 주파수 샘플은

역변환과 합성 윈도우를 통해 N 개의 시간 샘플로 복원된다. 그림 3 에서 제시된 방식으로 원본 시간 영역 신호를 복원하기 위해서는 그림 4 에서 제시된 것과 같은 행렬이 단위행렬이 되어야 한다. 이를 만족시켜 주는 변환 공식은 식 4 와 같다.

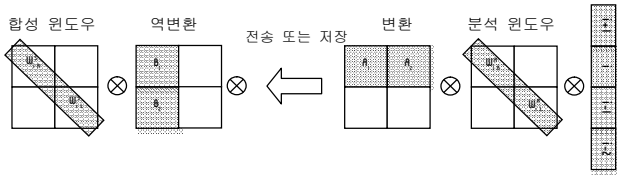

그림 3. MDCT 를 이용한 블록 변환에서 i 번째 프레임에 대한 처리 예. 회색 영역은 0 이 아닌 값을 가짐.

$$X_i[k] = \sum_{n=0}^{N-1} w_a[n] x_i[n] \cos\left(\frac{2\pi}{N}(n+n_0)\left(k+\frac{1}{2}\right)\right) \qquad \text{for } k=0,\ldots,N/2-1$$

$$x_i'[n] = w_s[n]\frac{4}{N}\sum_{k=0}^{N/2-1} X_i[k]\cos\left(\frac{2\pi}{N}(n+n_0)(k+\frac{1}{2})\right) \qquad \text{for } n=0,\ldots,N-1 \qquad (4)$$

상기 시간-주파수영역 변환 방식 및 이와 유사한 다양한 주파수 영역으로의 변환 방식은 추후 심리음향모델을 통해 얻어진 마스킹 커브를 이용하여 동적 비트 할당을 받음으로써 양자화시 데이터 압축 효과를 얻을 수 있다.

3. 차원 축소

차원 축소는 $X^{(N\times M)}$행렬을 특정한 조건을 바탕으로 $A^{(N\times R)}$, $B^{(R\times M)}$ 두 개의 행렬의 곱으로 표현한다. 이 때, R 은 주로 N, M 보다 작은 값으로 지정이 되게 되며, 차원 축소 알고리즘 별 특정한 조건 및 R 의 크기에 따라 원본 행렬을 얼마나 잘 복원할 수 있는지가 결정된다. 차원 축소 알고리즘은 특정 행렬로 표현된 데이터를 좀 더 작은 데이터량으로 표현하는 것뿐만 아니라, 이 과정에서 데이터에 내재되어 있는 추상적인 정보까지도 함축적으로 얻어낼 수 있기 때문에, 데이터의 압축에 좋은 성능을 보인다. 그림 5 는 이러한 차원 축소 과정에서의 행렬 곱을 표현한다.

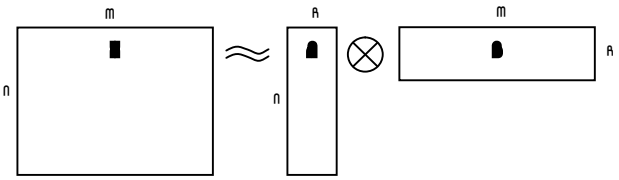

그림 5. 차원 축소 과정에서의 행렬 곱

차원 축소는 그 제한 조건과 입력의 특징에 따라 다양한 결과와 성능 차이를 보인다. 주요 알고리즘으로는, 주성분분석(PCA: Principal Component Analysis)[10] 독립성분분석(ICA: Independent Component Analysis)[11], 비음성 행렬 인수분해(NMF: Non-negative Matrix Factorization)[12] 등이 있다.

본 논문은 상기 대표적인 차원 축소 기술 이외에도, 그림 5 에서의 R 이 N 또는 M 보다 같거나 작은 상황에서 원본 행렬 X 를 A 와 B 의 내적으로 표현해주는 모든 차원 축소 방식을 대상으로 한다.

가. 주성분분석

주성분 분석은, 샘플 데이터의 공분산을 최대화하는 방향으로의 축변환을 통해, 데이터의 분포를 고려하면서 데이터를 전체적으로 대표

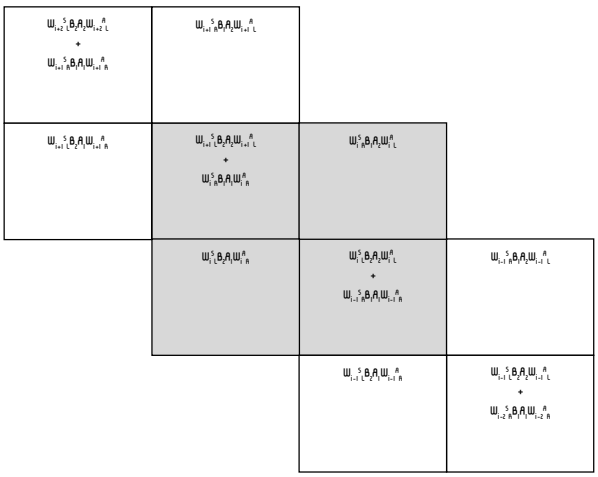

그림 4. Overlap-and-add 를 거친 후의 행렬 구조

할 수 있는 순서대로 주요한 방향 변환을 제공해준다. 그림 5 에서의 행렬 차원 축소 방식에서 해석하면, 주성분분석은 그 결과로써 A 의 열벡터가 행렬 X 를 축변환 해주는 위한 기저 벡터 역할로써 학습되게 된다.

나. 독립성분분석

독립성분분석은 전체 데이터 내에 독립적으로 존재하는 확률 분포를 학습함으로써, 직교하는 기저 벡터만을 도출할 수 있는 주성분분석에 비해 유연한 분포 대응이 가능하다. 일반적으로 독립성분분석은 주성분분석에 비해 좀 더 좋은 데이터 압축 성능을 보인다고 알려져 있으며, 특히 음원분리(Blind Source Separation)에서 독립적인 확률분포를 가지는 음원을 분리해내는 데에 좋은 성능을 보이고 있다. 또한 독립성분분석을 통한 데이터 압축이 심리음향모델을 통한 양자화 결과보다 좋은 음질을 보인다는 연구 또한 독립성분분석의 데이터 압축 성능을 단적으로 보여준다. [13]

다. 비음성 행렬 인수분해

비음성 행렬 인수분해는 음이 아닌 행렬에 대해 내적 연산을 통한 업데이트 방식을 통해 비음성을 보존하는 행렬 인수분해를 학습한다. 비음성 행렬 인수분해는 기존 차원 축소 방식들에 비해, 비음성 행렬만을 대상으로 한다는 점에서 한계점이 있지만, 학습 결과 행렬 A, B 가 기존 방식들에 비해 더 희소한 특징을 가지고 있으며, 이는 인간의 뇌가 데이터를 처리할 때 뇌의 특정 부분만 반응함으로써 희소한 표현을 한다는 점에서 뇌의 특성을 더 잘 반영함과 동시에 더 나은 데이터 압축 성능을 보여준다.

4. 스케일을 이용한 부분 차원 축소

주파수 영역으로 변환된 신호는 일반적으로 주파수 밴드의 개수(N)×전체 프레임의 개수(M)인 행렬로 볼 수 있다. 이 전체 행렬을 원본 X 로 간주하고, 차원 축소를 수행하면, R 에 따라 점진적으로 압축률이 다른 압축 결과를 얻을 수 있다. 그러나 이 경우, 행렬 전체를 입력 신호로 보아야 하기 때문에, 실시간 부호화가 불가능하며, 신호 전체 길이만큼의 부호화 지연이 발생한다. 이를 해소하기 위해, 본 논문에서는 특정 개수만큼의 프레임을 묶어서 여러 번의 차원 축소를 수행하는 방식을 제안한다.

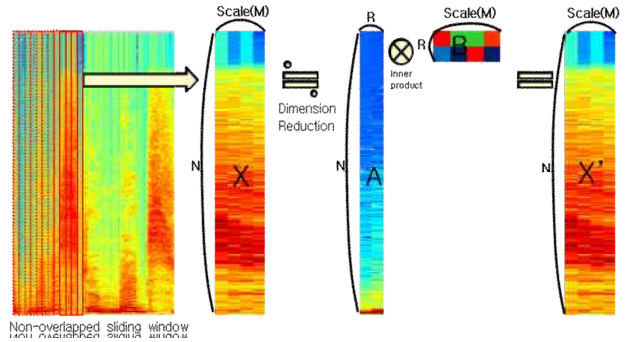

그림 6. 주파수 영역 신호의 차원 축소와 복원 과정

그림 6 은 주파수 영역으로 변환된 신호에 대해 부분적/순차적으로 차원 축소를 수행하는 방식에 대해 설명하고 있다. 먼저 주파수 영역으로 변환된 신호를 하나의 행렬로 보았을 때, 그 중 시간적인 순서에 따라 미리 정해진 스케일 M 만큼의 프레임들만을 취하여 N×M 입력 행렬 X 로 간주한다. 그림 6 에서는 스케일 계수 M 이 4 이고, R 이 2 인 경우에 대한 예를 보여주고 있다. 입력 행렬 X 는 차원 축소 방식을 거쳐, N×R 행렬 A 와, R×M 행렬 B 의 내적으로 표현되며, 이 경우, 4N 개의 원본 데이터 샘플은 차원 축소에 의해 2N+8 개의 샘플로 압축되며, 이 샘플은 간단한 내적으로 원본에 대한 근사치를 추정한다. 원본 행렬과 근사 행렬의 차이, 즉 압축 후의 음질 저하는 R 과 M 의 비율, R/M 과 관계가 있으며, 차원 축소 알고리즘의 종류에 따라 그 성능에 차이가 있다.

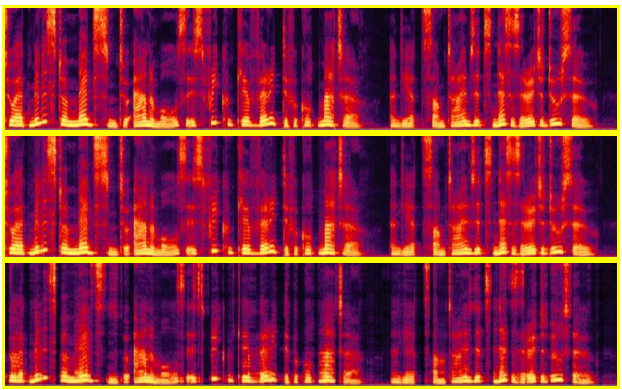
그림 7. 차원 축소를 이용한 음성 신호 압축 예 (위로부터 원본, 비음성 행렬 인수분해 사용, 주성분분석 사용. M=8, R=2 인 경우)

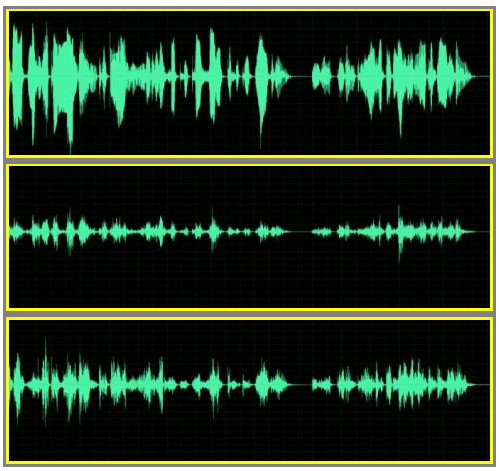
그림 8. 원 신호(음성)와 잔차 신호의 비교 (위로부터 원본, 비음성 행렬 인수분해 사용 후의 잔차 신호, 주성분분석 사용 후의 잔차 신호. M=8, R=2 인 경우)

5. 실험 결과

그림 7 과 8 은 각각 DFT 를 통해 주파수 영역으로 변환된 음성 신호를 비음성 행렬 인수분해와 주성분분석을 통해 차원 축소한 뒤 내적을 통해 복원한 결과와, 복원 신호와 원 신호 사이의 잔차 신호를 나타낸다. 스케일 계수 M 이 8 이고, R 은 2 인 경우로 상당히 많은 압축률을 보임에도 불구하고 특히 비음성 행렬 인수분해를 이용한 경우의 복원 신호가 원본에 비해 큰 손상이 없음을 확인할 수 있다. 음성 신호 뿐 아니라, 오디오 신호에도 동일한 경향성을 보이며, R/M 의 비율이 높아질수록 음질이 더 향상되는 경향이 있다.

6. 결론

상기 스케일을 이용한 주파수 영역 신호의 부분 차원 축소 방식에서, 스케일 정도와 축소 정도(R)는 응용에 따라 유동적으로 결정될 수 있다. 경우에 따라, 연속되는 프레임의 주파수 특성이 크게 변하지 않는 경우는 큰 스케일 값(M)과 작은 축소 차원 수(R)를 적용하고, 주파수 특성이 자주 바뀌는 구간은 작은 스케일 값과 상대적으로 큰 축소 차원 수를 적용하는 동적인 방식으로 신호의 특성에 대응함으로써 압축 효율을 높일 수도 있다. 주파수 영역 신호의 특성에 따라 다양한 차원 축소 알고리즘의 적용이 가능하며, 신호의 특성과 알고리즘의 특성에 따라 성능의 차이가 있을 것으로 예측되므로 보다 다양한 실험이 필요하다. 잔차 신호의 효율적 부호화 방식에 대한 추가적인 연구로 부호화 이득을 증가시킬 수 있으며, 청각적인 의미가 상실되는 차원 축소를 통해 얻은 기저 행렬 심리음향모델의 적용 가능 여부 및 추가적인 양자화 방법 연구가 진행 중이다.

7. 참고문헌

[1] ISO/IEC 13818-3, Information Technology, "Generic coding of moving pictures and associated audio, Part 3: Audio", 1994-1997.
[2] ISO/IEC 13818-7, Information Technology, "Generic coding of moving pictures and associated audio, Part 7: Advanced Audio Coding", 1997.
[3] M. Bosi, R. E. Goldberg, "Introduction to Digital Audio Coding and Standards", Kluwer Academic Publishers, pp.75-147, 2003.
[4] H. J. Nussbaumer, "Pseudo-QMF Filter Bank", IBM Tech. Disclosure Bull., vol. 24, pp. 3081-3087, November 1981.
[5] J. H. Rothweiler, "Polyphase Quadrature Filters – A new Subband Coding Technique", International Conference IEEE ASSP, Boston, pp. 1280-1283, 1983.
[6] ISO/IEC 11172-3, Information Technology, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s, Part 3: Audio", 1993.
[7] L. D. Fielder, M. Bosi, G. A. Davidson, M. Davis, C. Todd, and S. Vernon, "AC-2 and AC-3: Low Complexity Transform-Based Audio Coding," In N. Gielchrist and C. Grewin (ed.), Collected Papers on Digital Audio Bit-Rate Reduction, pp. 54-72, AES 1996.
[8] D. Sinha, J. D. Johnston, S. Dorward and S. R. Quackenbush, "The Perceptual Audio Coder (PAC)", in the Digital Signal Processing Handbook, V. Madisetti and D. Williams (ed.), CRC Press, pp. 42.1-42.18, 1998
[9] J. P. Princen, A. Johnson and A. B. Bradley, "Subband/Transform Coding Using Filter Bank Designs Based on Time Domain Aliasing Cancellation", Proc. Of the ICASSP, pp. 2161-2164, 1987
[10] I. T. Jolliffe, "Principal Component Analysis", Springer, 2002.
[11] A. HyvÄarinen, J. Karhunen, E. Oja: Independent Component Analysis. John Wiley & Sons, Inc. (2001)
[12] D. D. Lee and H. S. Seung: Learning the parts of objects by non-negative matrix factorization. Nature 401 (1999) 788.791
[13] A. Ben-Shalom, S. Dubnov and M. Werman , "Improved Low bit-rate audio compression using reduced rank ICA instead of psychoacoustic modeling", IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP2003