한국음향학회지 제32권 제3호 pp. 269~278 (2013)
The Journal of the Acoustical Society of Korea Vol.32, No.3 (2013)
http://dx.doi.org/10.7776/ASK.2013.32.3.269

pISSN : 1225-4428
eISSN : 2287-3775

국가기록원 음성 기록물의 복원과 분석

Restoration for Speech Records Managed by the National Archives of Korea

오세진†, 강홍구

(Sejin Oh and Hong-Goo Kang)

연세대학교 전기전자공학과

(접수일자: 2012년 12월 3일; 채택일자: 2012년 12월 28일)

초 록: 국가기록원의 음성 기록물은 우리나라의 근현대사를 담은 중요한 기록물이다. 하지만 아날로그로 녹음된 방식은 시간이 지남에 따라 손실을 피할 수 없어 디지털로 변환하여 관리 및 서비스할 필요성이 있다. 그에 따라 왜곡이 발생한 부분에 대해 본래의 정보를 복원하는 작업은 매우 중요하며, 본 논문은 음성 기록물의 훼손 종류에 따라 4가지의 카테고리로 분류하고 음량, 정상 잡음, 돌발 잡음에 맞는 복원 알고리즘을 적용하였다. 그 결과 음량은 음성 존재구간에 대해서 -26 dBov로 조정했고 SNR은 10 dB이상 상승하였다. 특히 기존에는 음성이 훼손된 부분을 순차적으로 청취하여 개별적으로 문제를 해결해야 했기 때문에 방대한 자료를 복원하기는 불가능 했지만 자동 복원 알고리즘을 도입하여 보다 효율적인 방식으로 복원할 수 있게 되었다.
핵심용어: 음성, 기록물, 잡음, 정상, 돌발, 복원

ABSTRACT: The speech recording of the National Archives of Korea contains very important traces which represent modern times of Korea. But the way to be recorded by analogue is easily contaminated as time goes by. So it has to be digitalized for management and services. Consequently, restoration method of distorted speech is needed. We propose the four classes for each distortion kind and apply restoration algorithms for the cases of speech level, stationary noise and abrupt noise. As a result, speech volume adjusts to -26 dBov for only on the speech region and SNR improves above 10dB. Especially, conventional way to remove the noise is almost impossible because we need to listen to all of them but it can be more effective by adaptation of auto restoration algorithm.
Keywords: Speech signal processing, Archives, Noise, Stationary, Abrupt, Impulsive
PACS numbers: 43.72. Ar

I. 서 론

지난 백여 년간 음성 및 오디오 데이터를 효과적으로 저장하기 위한 기술은 매우 급격히 발전되어 왔다. 특히, 디지털 방식으로 저장된 데이터는 아날로그 방식과는 달리 시간에 따른 열화 없이 반영구적으로 보존할 수 있으므로 그만큼 보관 및 관리하는데 드는 노력이 현저하게 줄어들며 언제든 처음 녹음 했을 때의 음질을 유지할 수 있다. 하지만 아날로그 방식은 시간이 지남에 따라 녹음된 LP판이나 자기테이프의 변형에 따른 데이터의 손실을 피할 수 없으며, 반복적으로 재생 할 경우에도 저장 매체의 변형이 따를 수밖에 없는 한계점을 지니고 있다.

이러한 문제를 해결하기 위해 현재 아날로그 데이터를 디지털로 변환하기 위한 필요성이 대두되고 있으며, 변환 과정에서 왜곡이 발생한 부분에 대해 신호처리 기법을 이용하여 본래의 정보를 복원하려는 작업 역시 매우 중요하다. 예를 들면 LP판의 경우 판 위에 위치해 있는 핀이 비정형적으로 동작함에 따라 임펄스 형태의 잡음을 생성하며, 자기 테이프 또한

†**Corresponding author:** Sejin Oh (vivid@dsp.yonsei.ac.kr)
Yonsei University B601 Engineering Building 134 Sinchon-dong, Seodaemun-gu, Seoul 120-749, Republic of Korea.
(Tel: 82-2-2123-4534, Fax: 82-2-364-4870)

먼지나 이물질, 그리고 테이프의 물리적 변형에 의해 다양한 형태의 잡음이 생성된다.

iZotopeRX는 오디오 신호를 복원하는 프로그램으로 사용자로 하여금 직접 잡음의 위치와 종류를 식별하게 하고, 이후 복원 기술을 적용하는 반자동 형태를 취하고 있다.[1] 이 때문에 방대한 음성 자료를 순차적으로 청취하여 훼손된 부분을 판별한 후, 개별적으로 문제를 해결하는 것은 시간 및 비용 측면에서 매우 비효율적이며, 그 효과 또한 기대하기 어렵다. 따라서 자동으로 왜곡을 보정하고 복원하기 위한 연구에 대한 중요성은 매우 크다. 자동 복원 시스템의 효율을 높이기 위해서는 왜곡의 특성에 따라 신호를 분류하기 위한 기술이 선행되어야 하며, 신호처리 기법을 응용하여 각각의 왜곡에 적합한 복원 방식을 개발하여야 한다.

국가기록원 음성자료는 우리나라의 근현대부터 현재까지의 중요하고 의미 있는 음성 자료들을 모아 놓은 것으로 이것을 분석하고 복원하는 것은 매우 큰 의의가 있다. 기존의 연구들은 특정한 잡음에 대한 실험을 하였으나,[16,17] 본 논문은 축적된 매우 방대한 양의 데이터베이스에 대한 훼손 정도 분석과 복원을 한다.

II. 국가기록원 음성자료

실험에 사용한 데이터베이스는 국가기록원에서 현재 소장하고 있는 데이터로, 15,735개의 음성 파일이며 전체가 약 16000시간의 재생 시간을 가지고 있는 방대한 데이터베이스이다. 대부분이 음성만을 포함하고 있는 것이 특징이다.

Fig. 1은 데이터베이스의 연도별 파일의 개수를 나타낸다. 1940년대에 녹음된 파일부터 최근 2008년까지 녹음된 파일이며 주로 1990년대 녹음된 데이터가 주를 이루고 있다. 녹음 매체별 파일의 개수는 Table 1과 같다.

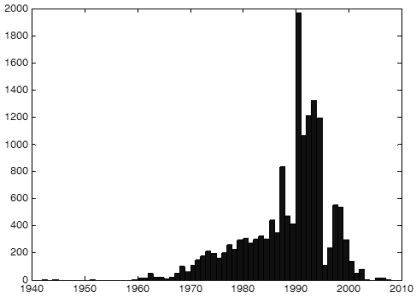


Fig. 1. Number of files for each year.

Table 1. Number of files for recording media.

cassette tape	reel tape	CD
12413	2025	730

1900년대에 주로 사용되었던 녹음테이프의 숫자가 가장 많은 빈도수를 차지하고 있다.

III. 음성의 분류 및 복원 알고리즘

3.1 음성 자료의 분류

Fig. 2와 같이 음성 자료를 구분하기 위해서 4가지의 파라미터를 사용하며 왼쪽부터 차례대로 잡음을 검출한 뒤에 복원하는 순서를 거치게 된다. 우선 음압을 측정하여 음량의 크기가 너무 작거나 큰 경우는 음량 카테고리 포함시킨다. 음량은 -26 dBov로 조정하였다. 잡음 카테고리는 정상(stationary) 잡음과 돌발(abrupt) 잡음으로 구분할 수 있다. 정상 잡음은 배경 잡음의 통계적 특성에 변화가 별로 없는 신호로써 잡음의 파워 스펙트럼을 추정하여 스펙트럼 신호 대 잡음비(SNR)를 예측하는 방식으로 잡음이 섞여있는 정도를 측정할 수 있다. 돌발 잡음은 단구간 에너지의 2차 미분 계수와 고대역 에너지의 양을 측정하여 검출하고 이것을 선형예측기법을 사용하여 품질을 개선한다. 마지막으로 음성 소실 카테고리는 음성의 정보가 크게 소실된 경우로서 주파수 밴드별 에너지를 기준으로 측정하여 판단할 수 있으며, 이 경우에는 복원이 매우 어렵기 때문에 구분만 하도록 한다.

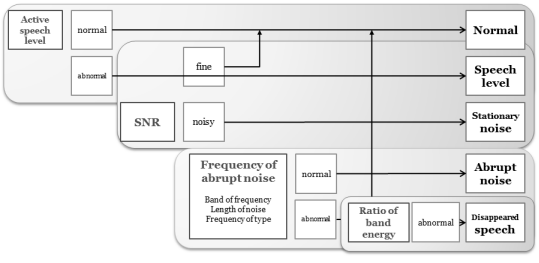


Fig. 2. Classification method of speech and category.

3.2 음량 측정 및 조절

음량을 측정할 때에는 전체 구간이 아닌 음성 구간만을 기준으로 음량을 측정해야 한다. 이것을 유효 음압 수준(Active Speech Level)이라고 한다.[11]

우선, 음성 신호의 envelope, q_i를 구하기 위해 다음과 같은 식을 이용한다.

$$p_i = gp_{i-1} + (1-g)|x_i|,$$
$$q_i = gq_{i-1} + (1-g)|p_i|. \quad (1)$$

여기서, q_i는 음성을 exponential 평균한 값이다. g는 시간 상수의 값으로 $g = e^{-t/T}$로 정의 된다. 여기서 T는 시간 상수로 0.03초 값을 쓰고 t는 샘플간의 시간 차로 샘플링 주파수의 역수 값이다.

음성 신호의 envelope을 임계값과 비교해서 크면 음성이 임계값보다 활성화 되었다고 하고 그 때마다 숫자를 누적한다. 각각의 샘플마다 값이 지속적으로 누적되면 묵음(혹은 정상 잡음)이 있는 부분에서의 값은 매우 커지고 상대적으로 큰 값을 가지는 음성 부분의 값은 묵음 구간에 비해 누적 량이 적다. 이 후 유효 수준의 파워와 임계값의 파워가 margin M (=15.9 dB, tolerance 0.5) 안에 들어오게 되면 그 값을 유효 음압 수준으로 정의한다.[11] 그리고 이후 음량 조절 시 이득값을 곱해서 −26 dBov로 일정하게 음량을 조절한다.

3.3 신호 대 잡음비 측정과 정상 잡음 제거

음성 신호는 정상 잡음이 더해진 형태로 나타나기 때문에 음성의 크기가 조절 된 이후에는 정상 잡음의 크기 또한 함께 변화하는 문제가 있다. 이 논문에서는 정상 잡음의 제거를 위해 MMSE 예측기를 사용한 OM-LSA(Optimally Modified Log-Spectral Amplitude) 방식을 도입하였다.[2] 잡음 추정기는 음질 향상 시스템의 전체 성능을 결정하는 핵심부분으로 잡음이 음성에 비해 상대적으로 느리게 변화한다는 가정에 근거하여, 일반적으로 음성이 존재하지 않는 구간에서 측정된 신호의 평균 파워를 잡음의 파워 스펙트럼으로 추정한다. 여기서 계산된 잡음의 파워 스펙트럼 정보는 신호 대 잡음비 추정기에 사용되고 여기서 계산된 선행 신호 대 잡음비, 사후 신호 대 잡음비 등의 정보는 이득 추정기에서 사용된다. 이 때, 신호 대 잡음비 추정기에서 추정된 선행 신호 대 잡음비(a priori SNR)값을 가지고 정상 잡음이 얼마나 포함 되어 있는지를 측정하였다.

3.4 돌발 잡음 위치 검출

돌발 잡음의 경우 일반적인 잡음과 달리 그 크기가 매우 크고, 빠르게 변화하며 잡음이 존재하는 시간이 매우 짧은 특성을 갖는다.[3] 이와 같은 특성으로 인해 돌발 잡음의 크기를 추정하는 것은 매우 어려운 일이며, 특히 음성이 존재하는 구간에서 돌발 잡음이 발생하는 경우엔 그 크기를 추정하기 매우 어렵다. 따라서 일반적인 돌발 잡음 제거 기법은 비선형 필터를 사용하는 방향으로 개발되었다.[4-7] 하지만 충격 잡음이 존재하는 구간을 찾지 못하면 음성이 왜곡되는 문제가 있다.

신호의 단구간 에너지나 변화량을 관찰하여 돌발 잡음이 존재하는 구간을 결정할 수 있는데 이 때 일정 기준 값보다 큰 구간을 선택하게 된다. 이런 배경 신호를 추정하기 위해서는 추가적인 기법이 필요하다. 특히 피치의 특성이 시간 축에서는 충격잡음과 유사한 특성을 가지므로 Whitening에 대한 성능 향상도 기대할 수 없다.[8,9] 이를 해결하기 위해 모음의 주기성을 이용한 알고리즘도 제안되었으나 충격잡음이 반복해서 나타나는 경우에서 취약하다.[10]

본 논문은 주파수 축에서는 고대역 에너지를 이용하고 시간 축에서는 2차 미분 계수의 단구간 에너지를 이용한 돌발 잡음 검출방법을 사용하였다.[12]

3.4.1 2차 미분 계수를 이용한 돌발 잡음 검출

2차 미분 계수는 신호가 급격하게 변하는 부분에서 큰 값을 가지게 되기 때문에 시간축에서 신호가 급변하는 부분을 찾아낼 수 있다는 장점을 가진다.[5] 입력 신호를 $x[n]$으로 정의하면 입력의 이차 미분 계수 $z[n]$은 다음과 같다.

$$z[n] = D^2x[n] = x[n-1] - 2x[n] + x[n+1]. \quad (2)$$

그에 대한 단구간 에너지 $w[n]$은 다음과 같다.

$$w[n]=\left[\frac{1}{N+1}\sum_{j=-N/2}^{N/2}z^2[n+j]\right]^{\frac{1}{2}}. \quad (3)$$

배경 신호를 추정하고 그보다 급격하게 값이 튀는 부분을 찾기 위해서 RMF(Recursive Median Filter)를 이용한다. 배경 신호의 2차 미분 계수의 단구간 에너지 $b[n]$은 다음과 같다.

$$b[n]=Med\begin{Bmatrix}b[n-M],\ldots,b[n-1],w[n],\\ w[n+1],\ldots,w[n+M]\end{Bmatrix}. \quad (4)$$

식에서 M은 필터의 좌, 우 길이를 나타낸다. 이 값을 상수 C_t와 비교하여 충격 잡음을 검출한다.

$$d[n]=\frac{|w[n]-b[n]|}{b[n]},\\ g_t[n]=\begin{cases}1, & d[n]>C_t\\ 0, & otherwise\end{cases}. \quad (5)$$

$g_t[n]$은 충격 잡음 검출 결과를 나타낸다. 여기서 C_t는 5로 설정하였다.

하지만 2차 미분 계수가 지니는 문제점은 음성의 유성음 구간에서 피치의 영향 때문에 False alarm이 일어난다는 것이다.

3.4.2 고대역 에너지를 이용한 돌발 잡음 검출

음성의 모음은 주파수의 저대역에 에너지가 집중되어 있는데 반해 돌발 잡음은 전체적으로 평탄한 주파수 응답을 갖는다.[4] 이 성질을 이용하여 고대역에 있는 에너지를 이용하여 돌발 잡음의 발생 위치를 추정할 수 있다.[12]

일반적으로 돌발 잡음은 고대역에 자리 잡고 있기 때문에 본 논문에서는 15 k에서 17 k대역의 에너지를 파라미터로 사용하였다. 물론 대부분의 음성 자료가 48 k나 혹은 44.1 k로 샘플링 되어 있어서 20 k 이상 대역의 정보도 얻을 수 있지만 카세트 테이프냐 혹은 릴 테이프냐에 따라서 최대 주파수가 다르기 때문에 17 k 이상의 대역은 사용하지 않았다. 고대역 에너지 $E_H(l)$은 다음과 같이 정의된다.

$$E_H(l)=\sum_{k=15kHz}^{17kHz}|X(k,l)|^2. \quad (6)$$

$X(k,l)$은 $x[n]$의 Fourier 변환 계수를 의미하며 k와 l은 각각 주파수와 프레임 인덱스이다. 이후로는 2차 미분 계수를 사용했을 때와 마찬가지로 RMF와 배경 신호에 대한 비를 고려한다.

$$\overline{E}(l)=Med\begin{Bmatrix}\overline{E_H}(l-30),\ldots,\overline{E_H}(n-1),\\ E_H(l),E_H(l+1),\ldots,E_H(l+30)\end{Bmatrix}. \quad (7)$$

배경신호 $\overline{E}(l)$로부터 정규화 과정을 거쳐 기준값과 비교한다.

$$R_H(l)=\frac{|E_H(l)-\overline{E_H}(l)|}{\overline{E_H}(l)},\\ g_H(l)=\begin{cases}1, & R_H(l)>C_f\\ 0, & otherwise\end{cases}. \quad (8)$$

$g_H(l)$은 주파수 축에서의 돌발 잡음 검출 결과를 나타낸다.

최종 돌발 잡음 구간은 고대역 에너지를 이용하여 잡음이 검출된 구간 안에서 2차 미분 계수 또한 돌발 잡음이어야 최종 결과를 돌발 잡음으로 결정한다.

3.5 음성 신호에 대한 모델링이 포함된 돌발 잡음 제거 시스템

일반적으로 음성 신호, 특히 모음은 formant와 pitch 정보로 나누어 모델링한다. 선형 예측 필터(LPC)의 잔여 신호에 장구간 예측 기법을 적용하면 pitch 정보를 모델링 할 수 있다.[9]

장구간 예측 기법에 의하면 현재의 잔여 신호는 한 pitch lag 이전의 잔여 신호에 일정한 pitch gain을 곱한 것으로 모델링 할 수 있다.

$$\tilde{r}(n+(l-1)M)=g_p(l)r(n+(l-1)M-\tau_p(l)),\\ n=0\sim M-1. \quad (9)$$

식에서 $r(n)$은 선형 예측 필터에 의한 잔여 신호를 나타내며 l은 장구간 예측 필터를 위한 프레임 인덱스, M은 프레임 길이를 나타낸다. 또한 $\tau_p(l)$과 $g_p(l)$은 pitch의 주기와 pitch gain으로써 아래와 같이 구할 수 있다.[13-14]

$$\tau_p(l) = \arg\max \frac{\sum_{n=0}^{M-1} r(n+(l-1)M)r(n+(l-1)M-\tau)}{\sqrt{\sum_{n=0}^{M-1} r^2(n+(l-1)M-\tau)}},$$
$$g_p(l) = \frac{\sum_{n=0}^{M-1} r(n+(l-1)M)r(n+(l-1)M-\tau_p)}{\sum_{n=0}^{M-1} r^2(n+(l-1)M-\tau_p)}. \quad (10)$$

pitch의 주기는 현재 프레임과의 상호 상관도가 가장 높은 지연 값 τ를 찾는 것으로 구하게 된다. pitch의 주기를 찾는 프레임 단위 M은 일반적으로 약 5 ms이다. 위의 식에서 피치의 주기를 자연수가 아닌 소수 값을 가지도록 하기 위해서 주어진 신호를 3배로 interpolation한 신호에 상관도를 구하여 더 정확하게 pitch 모델링을 하였다.[13]

이후 일반적인 음성 합성 과정에서는 먼저 장구간 예측 필터를 이용하여 pitch를 재합성하고 선형 예측 필터를 이용하여 formant 정보를 다시 합성하는 순서로 진행되며 장구간 예측 필터를 이용한 pitch의 합성은 재귀적인 합성 기법을 이용한다.

$$\hat{r}(n+(l-1)M) = \bar{e}(n+(l-1)M) + g_p(l)\hat{r}(n+(l-1)M-\tau_p(l)). \quad (11)$$

$\bar{e}$는 pitch를 모델링 하고 난 후의 추정 오차이다.

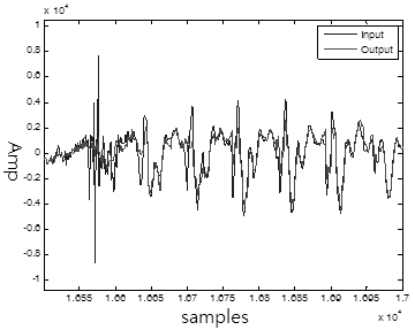


Fig. 3. Effect of recursive speech synthesis.

하지만 재귀적 합성 방법은 돌발 잡음이 median 필터에 의해 제거되면서 발생한 음성의 왜곡이 뒤에 pitch에 영향을 미치게 된다.

재귀적 합성 기법은 음성 부호화에서는 유용하지만 음성 신호 복원에는 적합하지 않다. 따라서 본 시스템에서는 신호의 잔여 신호 $\hat{r}$를 다음과 같이 구한다.

$$\hat{r}(n+(l-1)M) = \bar{e}(n+(l-1)M) + p(n+(l-1)M). \quad (12)$$

$p(n)$은 장구간 예측 필터를 이용해 모델링한 pitch 정보로써 돌발 잡음을 제거하기 전의 잔여 신호에서 pitch 모델링 오차를 뺀 값이다. 이와 같이 원 신호에서 pitch 신호를 모두 저장하였다가 돌발 잡음이 제거된 신호에 이를 더하면 왜곡이 이후 샘플에 영향을 미치지 않는다.[15]

Fig. 5는 장구간 예측 기법을 적용한 돌발 잡음 제거 시스템의 구조를 나타낸 것이다.

위의 돌발 잡음 제거 시스템은 앞선 음량 조절과 정상 잡음이 제거된 이후에 적용이 된다. 특히 정상 잡음을 제거 하고 나면 2차 미분 계수나 고대역 에너지의 배경 잡음이 제거 되면서 돌발 잡음 제거한 결과는 더욱 뛰어나다.[15]

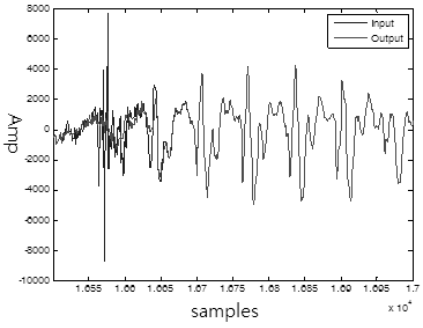


Fig. 4. Restored speech using pitch information.

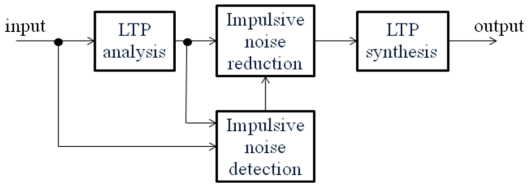


Fig. 5. Impulsive noise elimination using LTP.

3.6 소실 검출

국가기록원의 음성자료에서 소실 분류에 속하는 신호는 저대역에 신호가 몰려있고 음성의 특성이 나타나는 주파수 대역에는 정보가 전혀 나타나지 않는다.

소실 파일은 1 kHz아래 대역에 에너지가 몰려있는 특징을 가지고 있으며 이로 말미암아 음성이 웅웅거리는 소리만 나게 된다. 소실 신호는 음성 부분의 소리가 작지 않고 정상 잡음 또한 많이 포함되어 있지 않기 때문에 지금까지 설명한 방법으로는 구분해 낼 수 없는 문제점을 지니고 있다. 그렇기 때문에 전체 에너지와 1 kHz 아래의 저대역 밴드의 에너지의 비로 소실 음성을 검출해 내는 방법을 제안한다.

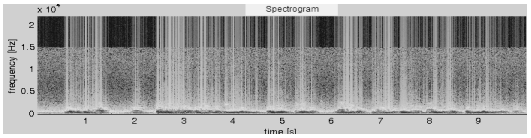
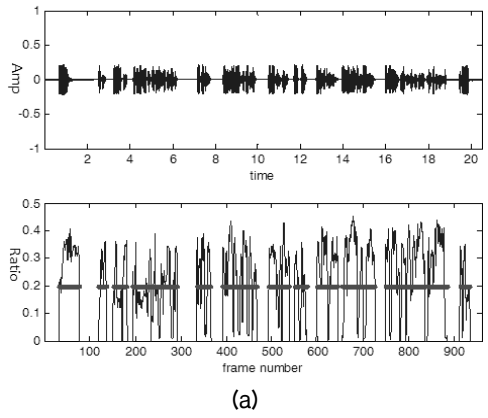
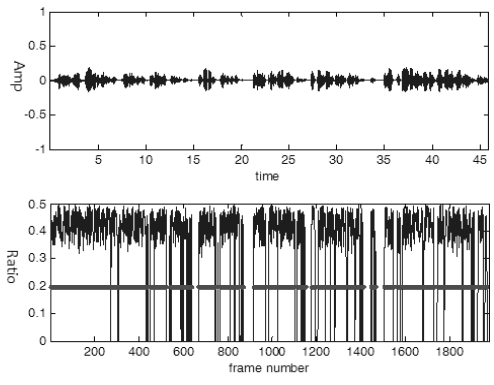


Fig. 6. Spectrogram of disappeared speech.



(a)



(b)

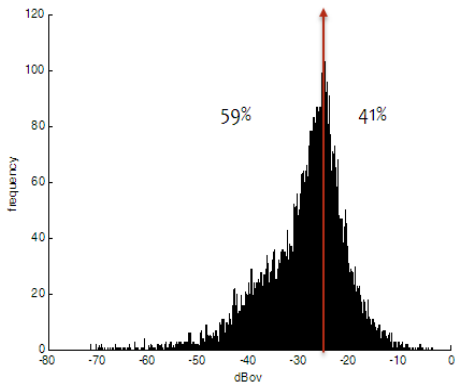
Fig. 7. Ratio of low band to overall energy and region of speech (thick solid line) for normal (a) disappeared (b).

Fig. 7의 상단 그래프에서 볼 수 있듯이 정상 신호와 소실 신호의 시간 축 그래프만으로는 두 가지를 구별할 수 없다. 하단 그래프는 해당 프레임의 에너지 비를 나타낸 것으로 정상 신호와 소실 신호의 경향이 차이가 나는 것을 확인할 수 있다. 굵은 선으로 표시된 부분이 음성이 있는 부분으로 이 부분에서 에너지 비의 평균을 보면 정상은 26%이고 소실은 45%로 소실이 매우 높은 것을 확인 할 수 있다. 소실 음성 자료들의 평균값은 44%로 기준 값 40%가 넘는 파일은 소실로 분류하였다.

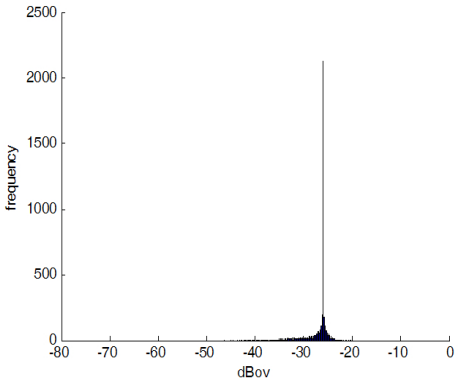
IV. 국가기록원 음성 자료의 분석 및 복원 결과

4.1 음량의 분포

Fig. 8은 음량의 분포를 나타내는 히스토그램이다. 복원 전 음량은 -26 dBov를 중심으로 가장 그 수가 많았지만, 음량이 작은 경우가 59%를 차지하여 녹음이

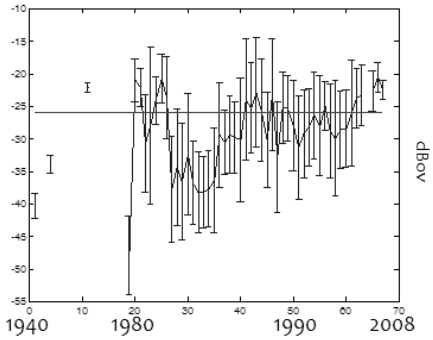


(a)

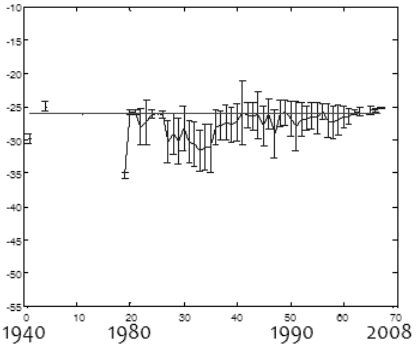


(b)

Fig. 8. Histogram of speech level before (a) after (b).

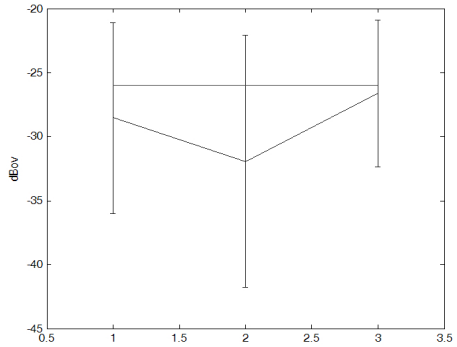


(a)

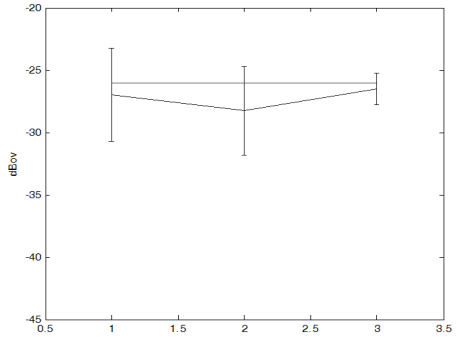


(b)

Fig. 9. Distribution of speech level for each year before (a) after (b).



(a)



(b)

Fig. 10. Distribution of speech level for each recording media before (a) after (b).

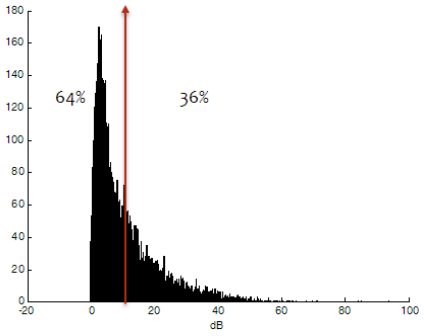
나 A/D 변환 과정에서의 에너지 레벨이 작게 변환되었음을 알 수 있다. 음량 조절 후에는 -26 dBov로 음량이 조절 된 것을 볼 수 있다. -26 dBov 외에 다른 값을 가지는 파일들은 복원 전에 음량이 너무 크거나 너무 작았던 탓에 음성이 있는 위치가 정확하지 않게 검출되었기 때문이다.

Fig. 9에서 위에 있는 그래프는 각각의 연도별로 음량의 평균과 그의 표준편차를 표시한 것이다. 연도별로 음량이 분포된 형태를 보면 1970년대 전후로 녹음된 자료들의 음량의 소리가 작은 것을 알 수 있다. 조절 후에는 -26 dBov를 중심으로 음량이 조절 되고 표준편차도 많이 줄어들었다.

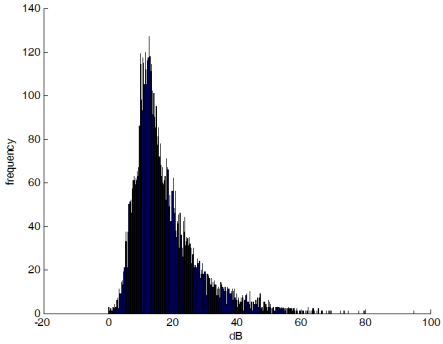
매체별로 음량의 분포를 살펴보면 릴로 녹음된 음성 자료들이 소리가 작으며 전체적으로도 음량이 작게 변환되어 있다. 조절 후에는 눈에 띄게 음량이 잘 조절 되었다.

4.1 신호 대 잡음비의 분포

복원 전 신호 대 잡음비의 분포를 보면 10 dB를 중



(a)



(b)

Fig. 11. Distribution of SNR before (a) after (b).

심으로 왼쪽에 치우쳐 있어 복원 필요성이 많음을 보여준다. 복원 후에는 전체적으로 10 dB 이상의 효

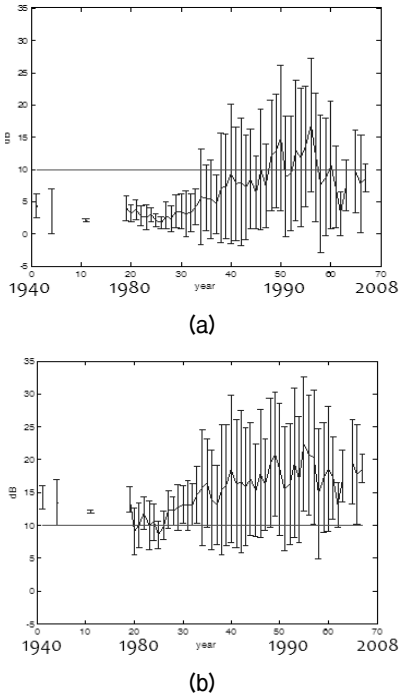


Fig. 12. Distribution of SNR for each year before (a) after (b).

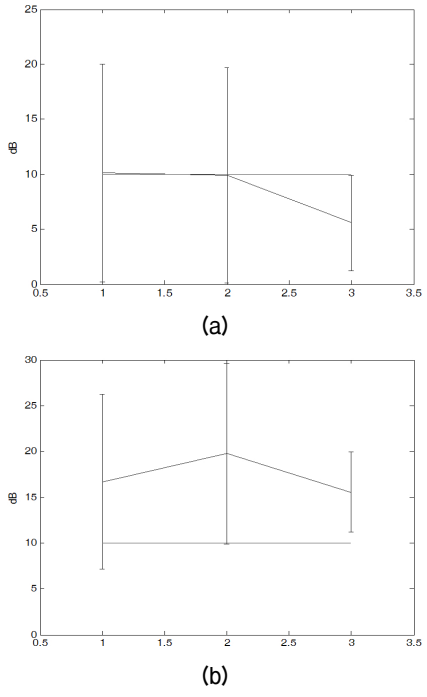


Fig. 13. Distribution of SNR for each recording media before (a) after (b).

과를 얻을 수 있었다.

Fig. 12는 신호 대 잡음비의 평균과 표준편차를 나타낸 그래프이다. 연도별 신호 대 잡음비를 보면 비교적 최근에 녹음된 파일들의 상태가 좋고 오래된 음성 자료일수록 녹음 상태가 좋지 않다는 것을 알 수 있다. 복원 후에는 전체적인 SNR이 10 dB의 선을 상회하게 나왔음을 알 수 있다.

매체별 신호 대 잡음비를 보면 카세트와 릴은 10 dB 평균값을 가지는 반면에 CD는 약간 낮은 값을 가지고 있음을 알 수 있다. CD로 녹음된 파일들의 경우 카세트와 릴에 비해 수가 적고 음악이 깔려있거나 혹은 음악만 있는 파일이 있어서 현재의 음성의 음량이나 신호 대 잡음비를 구하는 알고리즘에 적합지 않은 파일들이 일부 포함되어 있어서 작게 나온 경향을 띄었다. 복원 후에는 특히 릴에서의 성능이 매우 좋아졌다.

4.3 돌발 잡음의 검출 및 제거

Fig. 14의 왼쪽 그래프는 복원 전의 신호로 음량이 -33 dBov이고 SNR은 3 dB이다. 처음 부분과 중간에 돌발 잡음이 섞여 있다. 오른쪽의 그래프는 음량이 조절되고 정상 잡음이 제거된 뒤에 돌발 잡음을 제거한 결과이다. 돌발 잡음이 깨끗하게 제거되었을 뿐 아니라 소리도 매끄럽게 들리는 것을 확인할 수 있다.

돌발 잡음은 전체 파일에 대해서 검색을 해야 정확한 자료를 얻을 수 있지만 알고리즘을 모든 대상에 대해서 적용하면 시간이 많이 걸리기 때문에 처음에서 30초 떨어진 지점에서 10분 동안의 구간에서 돌발잡음을 검출하였다.

전체 파일 중에서 1,580개의 파일에서 돌발 잡음이 검출 되었으며 0.2초보다도 짧은 돌발 잡음들이 41%를 차지하고 있었다. 특정 구간에서만 찾은 것이기 때문에 이 자료는 참고 수치로 생각하는 것이 바람직하다.

4.4 소실 파일의 검출

소실 파일은 시작지점에서 30초 떨어진 지점에서 30초 동안의 구간에서의 에너지 비율을 측정하여 검

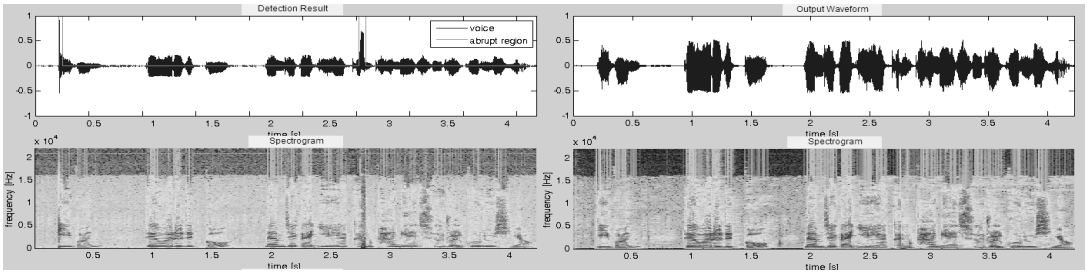


Fig. 14. Overall restoration result.

Table 2. Result of categorized speech-number of files and their percent.

speech level		stationary		abrupt		disappeared	
large	small	fine	noisy	normal	exist	normal	disap.
6455	9280	5671	10064	14155	1580	14836	899
41%	59%	36%	64%	90%	10%	94%	6%

출하였다. 전체 파일에서 899개가 소실 파일로 검출되었다.

4.5 국가기록원 음성자료의 구분 결과

Table 2는 전체 15,735개의 파일에 대한 결과이다.

IV. 결 론

국가기록원 음성 기록물은 우리나라의 근현대사를 보존하는 기록물로 매우 중요한 성격을 띤다. 본 논문은 훼손된 음성 기록물의 신호 특성을 기준으로 크게 네 가지의 카테고리를 만들고, 각각의 카테고리에 맞는 검출 방법을 이용하여 구분하였다. 또한 음량, 정상 잡음, 돌발 잡음을 복원하였으며 이는 앞으로 음성 기록물을 관리하고 서비스 하는데 도움이 될 것이다.

더 나아가 음성이 외부적으로 더해진 정상 잡음이나 돌발 잡음으로 침해 된 것이 아니라 소리 자체가 변형되어 왜곡된 경우를 복원하는 노력 또한 필요하다.

감사의 글

이 논문은 행정안전부 국가기록원 재원으로 2012년 기록보존기술 연구개발사업의 지원을 받아 수행된 연구임.

참 고 문 헌

1. iZotopeRX, audio repair toolkit http://www.izotope.com/products/audio/rx/.
2. Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator,” *IEEE Trans. on Acoust., Speech and Signal Process.* **33,** 443-445 (1985).
3. I. Cohen and B. Berdugo, “Speech enhancement for nonstationary noise environments,” *Signal Process.*, **81**, 2401-4218, (2001).
4. S. V. Vaseghi, *Advanced digital signal processing and noise reduction,* 2nd ed. *(*John Wiley & Sons, UK, 2000).
5. T. Kasparis and J. Lane, “Suppression of impulsive disturbances from audio signals,” *Electronics letters,* **29**, 1926-1927 (1993).
6. A. J. Efron and H. Jeen, “Detection in impulsive noise based on robust whitening,” *IEEE Trans. on Signal Process.***42**, 1572-1576 (1994).
7. S. R. Kim and A. Efron, “Adaptive robust impulse noise filtering,” *IEEE Trans. on Signal Process.***43**, 1855-1866 (1995).
8. I. Kauppinen, “Methods for detecting impulsive noise in speech and audio signals,” *in Proc. IEEE Int Conf. on Digital Signal Process.***2**, 967-970 (2002).
9. T. F. Quatieri, *Discrete-time speech signal processing,* (Prentice Hall, New Jersey, 2001).
10. J. Beh, K. Kim and H. Ko, “Noise estimation for robust speech enhancement in transient noise environment,” *in Proc. KSCSP 2007,* **24**, 35-36 (2007).
11. ITU-T, *ITU-T recommendation P. 56*, ITU-T, 2011.
12. M. S. Kim, H. S. Sin, H. G. Kang, “Time-Frequency Domain Impulsive Noise Detection System in Speech Signal” (in Korean), J. Acoust. Soc. Kr. Suppl. 2(s) **30**, 73-79 (2011).
13. ITU-T, *ITU-T recommendation G. 729*, ITU-T, 1996.
14. A. M. Kondoz, *Digital speech-coding for low bit rate communication systems*, (John wiley & sons, England, 1994).
15. M. Choi and H. Kang, “Transient noise reduction in speech signal with a modified long-term predictor,” EURASIP Journal on Advances in Signal Processing (2011).

16. Y. H. Son, Y. S. Park, H. S. Ahn, S. M. Lee, "An Improved Speech Absence Probability Estimation based on Environmental Noise Classification" (in Korean), J. Acoust. Soc. Kr. Suppl. 7(s) **30**, 383-389 (2011).
17. Y. G. Kim, H. J. Song, H. S. Kim, "Simultaneous Speaker and Environment Adaptation by Environment Clustering in Various Noise Environments" (in Korean), J. Acoust. Soc. Kr. Suppl. 6(s) **28**, 566-571 (2009).

저자 약력

▸ 오 세 진 (Sejin Oh)

2009년: 연세대학교 전기전자공학과 학사
2010년~현재: 연세대학교 전기전자공학과 석박통합과정 재학 중

▸ 강 홍 구 (Hong-Goo Kang)

1989년 2월: 연세대학교 전기전자공학과 학사
1991년 2월: 연세대학교 전기전자공학과 석사
1995년 8월: 연세대학교 전기전자공학과 박사
1996년~2002년: Senior Technical Staff Member, AT&T Labs-Research
2002년~2005년: 연세대학교 전기전자공학과 조교수
2005년~2011년: 연세대학교 전기전자공학과 부교수
2011년 9월~현재: 연세대학교 전기전자공학과 정교수