

2중발성환경에서 결합상관결수를 리용한 음향반향제거

리철의, 엄철남

경애하는 김정은동지께서는 다음과 같이 말씀하시였다.

《과학기술이 급속히 발전하고 그 갱신주기가 짧아지고있는 오늘의 현실적요구에 맞게 최신과학기술을 더 빨리 체득하고 그에 정통하기 위하여 피라게 노력하여야 합니다.》

화상회의체계에서 반향을 제거하는 문제는 회의에서 송수신되는 음성의 질을 보장하기 위한 중요한 요구로 나선다.

론문에서는 근단음성과 원단음성이 동시에 존재하는 2중발성환경에서 반향을 효과적으로 제거하기 위한 한가지 방법을 제안하였다.

1. 문 제 설 정

통신기술이 급속히 발전하면서 화상회의체계와 같은 음성통신체계들에서 음질을 보다 높이는 문제가 중요하게 제기되고있지만 회의과정에 발생하는 반향을 제거하고 음질을 보장하는것은 결코 쉬운 문제가 아니다. 이로부터 통신체계들에서는 효과적인 반향제거기술을 필수적으로 요구하고있다.

그림 1에 반향제거체계의 구성도[1, 2]를 보여주었다.

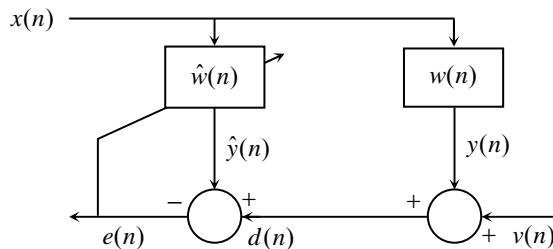


그림 1. 반향제거체계의 구성도

그림 1에서 원단신호 $x(n)$ 은 경로 $w(n)$ 을 통과하면서 반향 $y(n)$ 을 만든다. $v(n)$ 은 마이크로로부터 가까운 거리에 위치한 근단발성자로부터 발생되는 목적하는 신호이다. (근단신호)

두 신호 $y(n)$ 과 $v(n)$ 의 합신호 $d(n)$ 은 마이크의 입력신호로 된다. 반향제거방법의 목적은 이 합신호 $d(n)$ 에 포함되어있는 반향신호 $y(n)$ 을 제거하여 근단신호 $v(n)$ 을 얻는것이다.

반향제거에 리용되는 적응리파기는 반향이 생기는 경로 $\hat{w}(n)$ 을 근사적으로 모형화하는 기능을 수행하며 $d(n)$ 에서 이 적응리파기의 출력 $\hat{y}(n)$ 을 덜어 오차신호 $e(n)$ 을 계산한다. 이 오차신호에는 나머지만향신호와 함께 근단신호도 포함되게 되며 이 값은

려파기결수들을 갱신하는데 리용된다.

근단발성자가 없을 때 즉 $v(n)=0$ 일 때에는 적응려파기가 수렴하게 되며 반향이 발생하는 경로를 정확히 추정하여 반향을 충분히 제거할수 있다. 그러나 근단신호와 원단신호가 동시에 반향제거체계에 입력되는 경우 근단신호 $v(n)$ 은 적응려파기에 비상관성 잡음신호로 작용하여 적응려파기가 발산하게 만든다.

따라서 근단신호의 존재를 효과적으로 검출하는것은 반향제거에서 실지 중요한 역할을 한다. 이 문제를 2중발성검출(DTD: Double-Talk Detection)문제[1]라고 한다.

화상회의에서 반향신호를 효과적으로 제거하자면 2중발성검출기술과 적응반향제거기술을 효과적으로 결합하여 리용하여야 한다.

2중발성검출방법들에는 근단신호와 원단신호사이의 진폭비교를 리용하는 게이겔(Geigel)방법과 신호포락에 기초한 방법, 간섭성에 기초한 방법 등이 있다.

가장 일반적인 게이겔 2중발성검출방법에서는 원단신호 $x(n)$ 에서 길이가 L 인 신호의 최대값과 마이크입력신호 $d(n)$ 사이의 진폭준위비교를 진행한다.

결정변수는 다음과 같다.

$$\xi_G(n) = \frac{\max\{|x(n)|, |x(n-1)|, \dots, |x(n-L+1)|\}}{|d(n)|} \quad (1)$$

2중발성상태는 $\xi_G(n) < T_G$ 일 때 나타난다. 여기서 T_G 는턱값이다. 게이겔방법은 추가적인 잡음에 민감하며 반향준위는 대체로 6dB이하이다.

본문에서는 분할블록주파수영역적응려파알고리즘과 개선된 2중발성검출알고리즘을 제안하였다. 여기서는 결합상관결수를 받아들여 당시의 음성상태 즉 2중발성환경인가 아닌가를 판단하고 주파수영역에서의 적응려파처리를 진행함으로써 2중발성환경에서 반향제거알고리즘의 성능을 향상시켰다.

2. 분할블록주파수영역적응려파

음성통신체계들에서 반향제거의 성능을 높이자면 반향 특히 지연이 긴 반향신호를 제거하여야 한다.

그러나 이미 나온 최소평균두제곱(LMS: Least Mean Square)적응려파알고리즘은 그 계산량이 려파기길이에 비례하고 수렴속도가 느린 결함이 있다.

이로부터 본문에서는 분할블록주파수영역적응려파알고리즘을 적용하여 지연이 긴 반향신호를 제거하였다.

분할블록주파수영역적응려파방법에서는 입력신호를 블록단위로 고속푸리에변환하여 주파수영역에서 처리하므로 계산량이 보다 적으면서도 잘 수렴하게 된다.

[알고리즘 1] 분할블록주파수영역적응려파알고리즘

입력: 원단신호 $x(n)$, 마이크입력신호 $d(n)$

출력: 려파결수 \hat{W}

결음 1 임의로 초기화한 N 차원려파기 w_0 을 시간영역에서 N/P 개의 부분으로 나눈다. 즉 매 부분은 P 개의 려파기결수들을 포함한다.

$$w_{0i} = [w_0(iP), \dots, w_0((i+1)P-1)]^T, i=0, 1, \dots, \frac{N}{P}-1 \quad (2)$$

다음 매 w_{0i} 에 령채우기를 진행하여 차원수를 $M(M \geq P)$ 으로 만든 다음 고속푸리

에 변환을 리용하여 주파수영역으로 변환한다.

$$\mathbf{W}_{0i} = \mathcal{F} \begin{bmatrix} \mathbf{w}_{0i} \\ \mathbf{0}_{M-P} \end{bmatrix} \quad (3)$$

여기서 \mathcal{F} 는 M 차고속푸리에변환(FFT)행렬을 나타낸다.

걸음 2 시간영역의 원단신호 $x(n)$ 을 길이가 L 인 블록단위로 토막화한다.

$$\mathbf{x}_m = [x(mL), \dots, x((m+1)L-1)]^T, m = 0, 1, 2, \dots \quad (4)$$

여기서 m 은 블록번호, 파라미터 L 은 블록길이이다.

걸음 3 블록신호 \mathbf{x}_m 을 포함하여 앞서 입력된 N 개 점에 해당하는 원단신호를 M 개 단위로 분할하고 매 부분들을 고속푸리에변환에 의하여 주파수영역으로 변환한다.

$$\mathbf{X}_{mi} = \text{diag} \left\{ \mathcal{F} \begin{bmatrix} x[(m+1)L-iP-M] \\ \vdots \\ x[(m+1)L-iP-1] \end{bmatrix} \right\}, i = 0, 1, \dots, \frac{N}{P}-1 \quad (5)$$

걸음 4 주파수영역에서 려파기를 적용한 후 매 토막의 결과신호들을 모두 더하여 합신호를 계산한다.

다음 합신호에 거꿀푸리에변환을 적용하여 신호 $\hat{\mathbf{y}}_m$ 을 추정한다. 이때 선형중첩연산의 결과에서 뒤로부터 L 개의 점들에 해당하는 신호가 원단신호의 m 번째 블록신호속에 들어있는 반향신호 $\hat{\mathbf{y}}_m$ 으로 된다.

$$\hat{\mathbf{y}}_m = \begin{pmatrix} \mathbf{0}_{M-L} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_L \end{pmatrix} \mathcal{F}^{-1} \sum_{i=0}^{N/P-1} \mathbf{W}_{mi}^H \mathbf{X}_{mi} \quad (6)$$

여기서 옷첩수 H 는 에르미트연산자이다.

걸음 5 $\hat{\mathbf{y}}_m$ 을 마이크입력신호 $\mathbf{d}_m = [d(mL+1), \dots, d(m(L+1)-1)]^T$ 에서 덜어 오차신호 \mathbf{e}_m 을 얻는다.

$$\mathbf{e}_m = \mathbf{d}_m - \hat{\mathbf{y}}_m \quad (7)$$

걸음 6 주파수영역에서 려파기갱신의 걸음값을 계산하고 오차신호 \mathbf{e}_m 에 따라 적응 려파기결수들을 갱신한다.

$$\hat{\mathbf{W}}_{m+1i} = \hat{\mathbf{W}}_{mi} + \mathcal{F} \mathbf{G} \mathcal{F}^{-1} \Pi_m \mathbf{X}_{mi}^H \mathbf{E}_m, i = 0, 1, \dots, \frac{N}{P}-1 \quad (8)$$

$$\Pi_m = \mu \left(\sum_{i=0}^{N/P-1} \mathbf{X}_{mi}^T \mathbf{X}_{mi} + \gamma \mathbf{I}_M \right)^{-1}, \mathbf{E}_m = \mathcal{F} \begin{bmatrix} \mathbf{0}_{M-L} \\ \mathbf{e}_m \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \mathbf{I}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{M-P} \end{bmatrix} \quad (9)$$

여기서 μ 는 려파기의 수렴속도와 안정성에 영향을 주는 걸음크기파라메타이며 γ 는 자료넘침을 피하기 위하여 첨부한 매우 작은 상수이다.

3. 결합상관결수에 기초한 반향제거

화상회의체계에서는 2중발성검출기술을 리용하여 당시의 음성상태를 검출하고 근단 음성이 있는 구간에서는 려파기의 갱신을 중지하는 방법으로 반향제거를 위한 적응려파기의 수렴성을 담보한다.[1]

론문에서는 신호들사이의 상관결수를 리용하여 단일발성과 2중발성상태를 식별하는

한가지 방법을 제안하였다.

2중발성검출알고리즘에서는 분할블록주파수영역적응리파기의 출력을 리용하여 주파수영역에서 원단블록신호, 근단블록신호와 오차블록신호사이의 교차상관값으로부터 결합상관결수를 계산한다. 이 결수는 당시의 음성상태를 표현하며 오차신호를 평활화하는 파라미터로 적용되어 최종출력신호를 얻는다.

[알고리즘 2] 결합상관결수에 기초한 반향제거알고리즘

입력: 근단블록신호 d_m , 원단블록신호 x_m

출력: 반향이 제거된 ef_m

결음 1 근단블록신호 d_m , 원단블록신호 x_m 과 오차블록신호 e_m 의 전력스펙트르밀도를 각각 계산하고 평활화를 진행한다.

$$\begin{cases} S_m^d = \lambda S_{m-1}^d + (1-\lambda) d_m^H \cdot d_m \\ S_m^x = \lambda S_{m-1}^x + (1-\lambda) x_m^H \cdot x_m \\ S_m^e = \lambda S_{m-1}^e + (1-\lambda) e_m^H \cdot e_m \end{cases} \quad (10)$$

여기서 S_m^d , S_m^x 과 S_m^e 은 각각 전력스펙트르밀도이고 λ 는 평활화결수이다.

결음 2 원단블록신호 x_m 과 근단블록신호 d_m 의 호상전력스펙트르밀도 S_m^{xd} 와 오차블록신호 e_m 과 근단블록신호 d_m 의 호상전력스펙트르밀도 S_m^{ed} 를 각각 계산한다.

$$\begin{cases} S_m^{xd} = \lambda S_{m-1}^{xd} + (1-\lambda) S_m^x \cdot (S_m^d)^* \\ S_m^{ed} = \lambda S_{m-1}^{ed} + (1-\lambda) S_m^e \cdot (S_m^d)^* \end{cases} \quad (11)$$

결음 3 원단블록신호 x_m 과 근단블록신호 d_m 의 상관결수 $Coh_{xd}(m)$, 오차블록신호 e_m 과 근단블록신호 d_m 의 상관결수 $Coh_{ed}(m)$ 을 각각 계산한다.

$$\begin{cases} Coh_{xd}(m) = \frac{S_m^{xd} \cdot (S_m^{xd})^*}{S_m^x \cdot (S_m^d)^*} \\ Coh_{ed}(m) = \frac{S_m^{ed} \cdot (S_m^{ed})^*}{S_m^e \cdot (S_m^d)^*} \end{cases} \quad (12)$$

여기서 *은 공액복소수를 표시한다.

결음 4 $1-Coh_{xd}(m)$ 과 $Coh_{ed}(m)$ 중에서 최소값을 결합상관결수 P_{Coh} 로 선택한다. 이 결수는 현재음성상태의 검출정확도를 최대화한다.

$$P_{Coh}(m) = \min(1-Coh_{xd}(m), Coh_{ed}(m)) \quad (13)$$

결음 5 제거함수를 다음과 같이 얻는다.

$$Supp(m) = P_{Coh}(m)^{\beta(m)} \quad (14)$$

여기서 $\beta(m)$ 은 제거정도를 나타내는 결수이다.

결음 6 오차블록신호 e_m 에 제거함수 $Supp(m)$ 을 적용하여 최종출력신호벡터 ef_m 을 얻는다.

$$ef_m = Supp(m)e_m \quad (15)$$

이상의 과정으로부터 제거함수는 결합상관결수 P_{Coh} 에 의존하며 이 결수값은 각이한 음성상태에서 오차신호의 제거정도를 특징짓는다는것을 알수 있다. 즉 원단발성자와 근단발성자가 동시에 말할 때 즉 2중발성 환경에서는 제거정도를 줄여 될수록 근단신호들

을 유지하며 근단발성자가 말을 안할 때에는 제거정도를 크게 하여 환경에 존재하는 반향신호들을 가능한껏 제거한다.

4. 실험 및 결과분석

제안된 알고리즘의 성능을 평가하기 위하여 Matlab를 리용하여 모의실험을 진행하였다.

실험에 리용된 시험자료는 조선어음성화일로서 표본화주파수는 16kHz이고 음성지속시간은 약 45s이다.

음성신호들의 시간영역에서의 파형을 그림 2에 보여주었다.

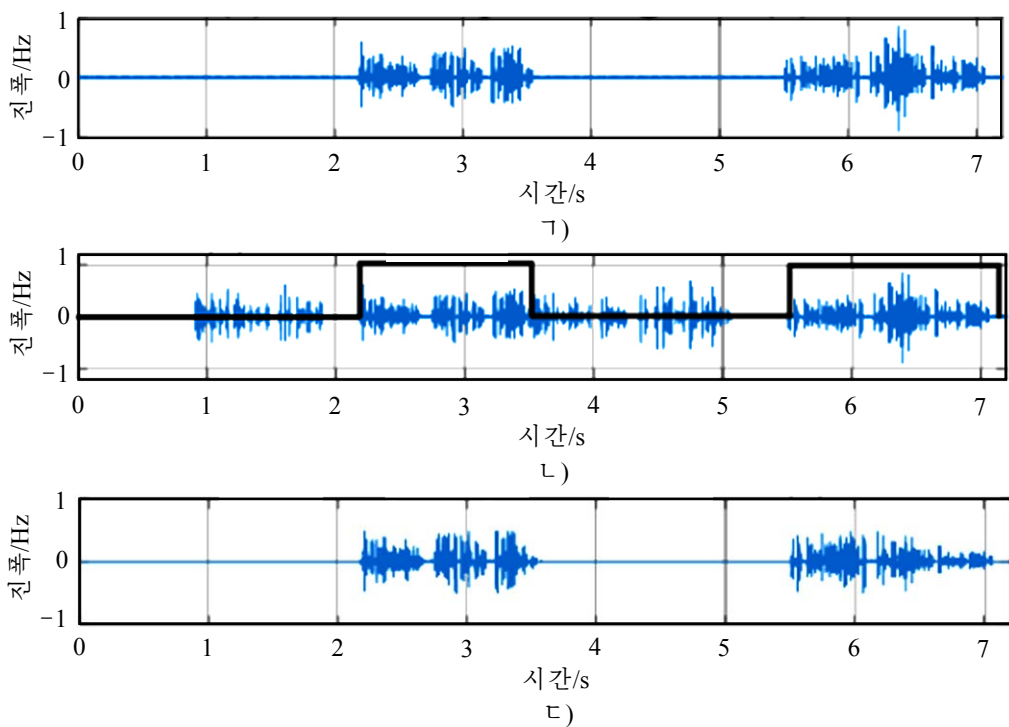


그림 2. 음성신호들의 시간영역에서의 파형

가) 근단신호 $v(n)$, 나) 마이크입력신호 $d(n)$, 다) 반향이 제거된 최종출력신호 $ef(n)$

그림 2의 나)에서 굵은 선은 가까운 근단신호를 표시하는 검출변수를 나타내므로 그것을 마이크입력신호에 중첩하여 표시한다.

그림 2로부터 개선된 알고리즘이 마이크입력신호에서 근단신호를 잘 보존하면서도 그속에 섞여있는 반향신호를 충분히 제거한다는것을 알수 있다.

다음으로 주관적인 음질평가를 진행하였는데 그 결과는 2중발성환경에서 반향제거체의 출력신호가 끊기거나 단어손실이 없이 음질이 개선된다는것을 보여준다.

반향제거알고리즘의 성능을 평가하기 위하여 이전의 NLMS, FLMS방법[1, 2]들과 제안된 알고리즘을 비교하였다.

불일치도[2]는 다음과 같이 정의된다.

$$MIS(dB) = 10 \log_{10} \left[\frac{\|w(m) - \hat{w}(m)\|^2}{\|w(m)\|^2} \right] \quad (16)$$

여기서 $\hat{w}(m)$ 은 추정된 적응러파기이며 $w(m)$ 은 실지반향경로이다.

2중발성검출알고리즘으로서 가장 일반적으로 리용되는 게이젤알고리즘과 제안된 결합상관결수에 기초한 2중발성검출알고리즘을 비교하였다.(그림 3) 비교를 진행하기 위하여 적응러파기의 반복수를 40 000으로 설정하고 15 000~20 000개의 표본을 가지는 3개의 음성신호자료를 리용하였다.

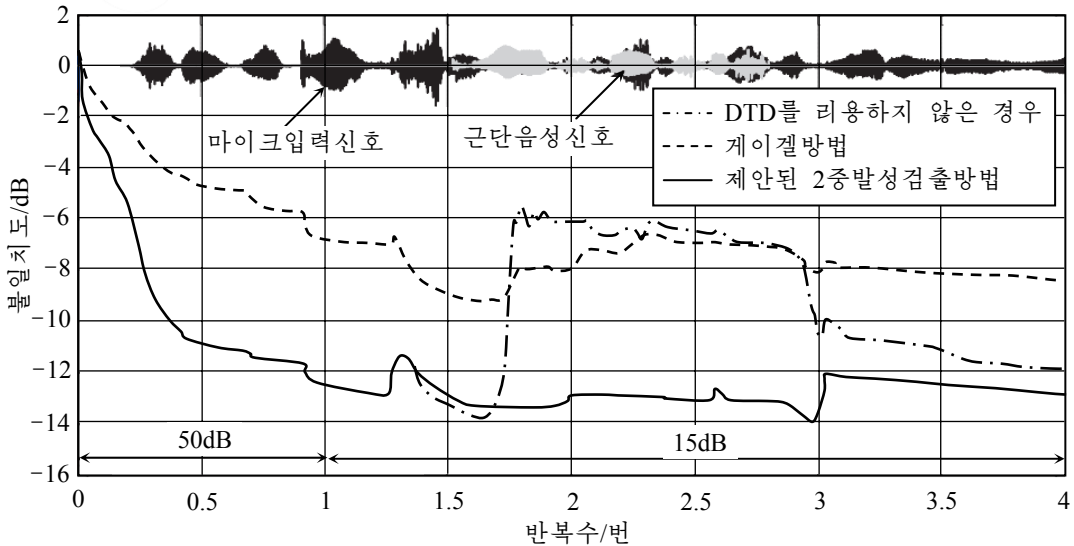


그림 3. 게이젤알고리즘과 제안된 결합상관결수에 기초한 2중발성검출알고리즘의 성능비교

그림 3을 통하여 제안된 결합상관결수에 기초한 2중발성검출알고리즘을 리용한 반향 제거성능이 2중발성검출기를 리용하지 않거나 게이젤알고리즘을 리용할 때보다 불일치도 값이 훨씬 작으면서 수렴속도가 빠르다는것을 알수 있다.

맺 는 말

분할블록주파수영역적응러파알고리즘과 제안된 결합상관결수에 기초한 2중발성검출알고리즘을 리용하여 음향반향제거체제를 구성하였으며 제안된 방법이 2중발성환경에서 음향반향을 효과적으로 제거하면서 높은 음질을 보장한다는것을 검증하였다.

참 고 문 헌

- [1] M. Fukui et al.; IEEE Transactions on Consumer Electronics, 60, 3, 468, 2014.
- [2] Y. Huang et al.; IEEE International Conference on Acoustics, Speech and Signal Processing, 596, 2017.

Acoustic Echo Cancellation Using Joint Correlation Coefficient in Double-Talk Situation

Ri Chol Ui, Om Chol Nam

In this paper, an improved double-talk detector algorithm is proposed to promote the performance of echo cancellation in the presence of speech issued from near-end speaker in video conference system.

Keywords: acoustic echo cancellation, double-talk detection