

# 텔레매틱스 시스템을 위한 반향제거 및 Barge-In 기능을 갖는 음성인터페이스

## Speech Interface with Echo Canceller and Barge-In Functionality for Telematic System

김 준\*, 배 건 성\*\*  
(Jun Kim\*, Keunsung Bae\*\*)

\*국방과학연구소 무인지술화 연구실, \*\*경북대학교 전자전기컴퓨터학부  
(접수일자: 2009년 4월 7일; 수정일자: 2009년 6월 1일; 채택일자: 2009년 6월 17일)

본 논문에서는 배경잡음과 반향이 존재하는 차량환경에서 음성인식 성능을 향상시키기 위해 상관계수를 이용한 동시통화 검출 알고리즘을 적용한 음향 반향제거기와 barge-in 기능을 갖는 음성 인터페이스를 구현하였다. 상관계수를 이용한 동시통화 검출 알고리즘은 입계치 설정 및 배경잡음의 영향 등으로 인해 검출 오류가 발생한다. 이를 보완하기 위해 동시통화 검출 조건으로 매 샘플마다 입력신호에서 추정된 배경잡음 및 반향신호의 평균 전력을 이용하여 동시통화 검출 오류를 줄였으며, 시변의 임계치를 적용한 후처리 단을 통해 시변의 잔여 잡음 성분을 제거하였다. 또한 안내음성 중에 음성입력이 가능하도록 barge-in 기능을 적용한 음성 인터페이스 시스템을 구현하였다. 제안한 음성 인터페이스 시스템은 동시통화 검출 오류와 이로 인해 발생하는 문제점을 효율적으로 해결할 수 있음을 실험을 통하여 확인하였다.

**핵심용어:** 음성인터페이스, 반향제거, Barge-in

**투고분야:** 음성처리 분야 (2,3)

In this paper, we develop a speech interface that has acoustic echo cancelling and barge-in functionalities in the car environment. In the echo canceller, DT (Double-Talk) detection algorithm using the correlation coefficients between reference and desired signals can make DT detection errors often in the background noise. We reduce the DT detection errors by using the average power of noise and echo estimated from the input signal. In addition, to make it possible for drivers to give speech command to the system by interrupting the speaker output, barge-in functionality is implemented with the combination of DT detection and appropriate gain control of the speaker output. Through the computer simulation with the assumed car environment and experiment in the real laboratory environment, implemented speech interface has shown good performance in removing acoustic echo signals in the noisy environment with proper operation of barge-in functionality.

**Keywords:** Speech interface, Echo canceller, Barge-in

**ASK subject classification:** Speech Signal Processing (2,3)

### I. 서론

오늘날 컴퓨터 및 정보통신기술의 발전으로 운행 중인 자동차에서도 GPS (Global Positioning System)를 통한 도로안내 및 관련 서비스 제공, 디지털 방송을 통한 교통 정보 제공 등과 같은 다양한 형태의 서비스를 운전자에게 제공해주는 텔레매틱스 (telematics) 시스템이 실용화

되고 있다 [1]. 새로이 도입되는 차량용 정보시스템들이 운전 중의 편리함을 추구하는 것이 목적이지만, 운행 중인 운전자의 시선유도나 손동작을 유발함으로써 운전의 위험성을 증가시킬 수 있다. 이러한 문제점을 해결하기 위해, 손과 눈이 자유롭게 못한 상황에서 안전 운전을 보장해 줄 수 있는 최적의 사용자 인터페이스로서 음성을 사용하는 HMI (Human Machine Interface) 기술에 대한 연구가 유럽 및 미국을 비롯한 선진국과 국내에서 활발히 진행되고 있다. 이러한 기술은 차세대 자동차의 경쟁력을 좌우할 필수적인 기술로 부각되면서 2001년 GM

책임저자: 김 준 (ksbae@mirbbs.knu.ac.kr)  
305-600 대전 유성우체국 사서함 35-1호 국방과학연구소 무인지술화 연구실  
(전화: 053-940-8627; 팩스: 053-940-8827)

의 OnStar를 시작으로 상업적인 시도가 국내외에서 본격적으로 시작되었다 [2]. 하지만, OnStar 시스템을 비롯한 초기의 음성인터페이스는 반향이나 잡음이 존재하는 환경에서의 인식 성능을 보장하지 못한다. 따라서 운행 중인 자동차와 같이 엔진소음 및 도로주행 소음 등 배경잡음이 많은 환경에서 음성인식 기술을 이용한 인터페이스를 제공하기 위해서는 다음과 같은 핵심기술의 개발이 필수적이다. 하나는, 운행 중인 자동차 환경에서 발생하는 잡음을 처리하는 기술 개발이며, 다른 하나는 텔레매틱스 시스템의 음성출력 도중에 운전자가 음성명령을 할 수 있도록 해주는 barge-in 기술 및 자동차 실내에서의 반향제거 기술이다. 이에 본 논문에서는 텔레매틱스 시스템에서 잡음에 강한 음성인식을 위한 전처리 단으로서, 반향제거기와 barge-in 기능을 갖는 음성인터페이스를 구현하였다. 논문의 구성은 다음과 같다. 먼저 2장에서는 차량환경에서 적용되는 일반적인 음성인터페이스와 기존 시스템이 가지는 문제점을 간략하게 서술한다. 3장에서는 이러한 문제점을 해결하고, 보다 강한 성능을 갖는 음성인터페이스를 구현하기 위해 본 연구에서 제안한 DT 시작점 및 끝점 검출 알고리즘, barge-in 기능 및 후처리 과정을 설명한다. 더불어 시뮬레이션 (simulation) 및 실시간 구현 실험 결과를 4장에서 제시하고, 마지막으로 5장에서 결론을 맺는다.

## II. 텔레매틱스 시스템을 위한 음성인터페이스

차량환경에서 음성인터페이스의 구현은 특히, 오디오 출력장치에 의한 반향신호의 영향을 크게 받게 되므로 음향 반향제거기는 반드시 필요하다. 스피커와 마이크 사이의 반향 경로가 변하지 않는다면 고정된 임펄스 응답을 측정하여 필터링 함으로써 반향을 제거할 수 있지만, 일반적으로 시간에 따라 통계적인 특성이 변하는 음향반향신호를 제거하기 위해서는 적응필터를 이용한 반향제거 기법을 이용한다. 이 방법은 기준이 되는 시스템의 출력신호로부터 시간에 따라 변하는 반향경로의 임펄스 응답을 추정하여 마이크로폰으로 입력되는 신호에서 추정된 반향신호를 제거하게 되며, 기본적인 구조는 그림 1과 같다.

$x(n)$ 은 음향시스템의 출력신호로서 적응필터의 기준신호 (reference signal)로 사용되고,  $y(n)$ 은 스피커에서

출력된 신호가 반향경로를 거쳐 마이크로폰으로 입력되는 반향신호를 의미한다.  $v(n)$ 은 사용자의 음성신호이며,  $d(n)$ 은 반향신호와 사용자의 음성신호 및 배경잡음  $n(n)$ 을 포함하는 마이크 입력신호이다.  $e(n)$ 은 적응필터의 추정 오차신호 (error signal)로서 반향제거기에서의 잔여 반향 (residual echo)을 의미한다. 적응필터는 기준신호  $x(n)$ 과 추정오차 신호  $e(n)$ 을 이용하여 스피커와 마이크 사이의 반향 경로를 추정함으로써  $\hat{y}(n)$ 을 계산하고, 반향신호  $y(n)$ 에서  $\hat{y}(n)$ 을 빼줌으로써 반향신호를 제거한다. FIR 필터로 구현되는 적응필터의 적응 알고리즘에는 LMS (Least Mean Square), NLMS (Normalized LMS), BNLMS (Block NLMS), LS (Least Square), RLS (Recursive LS) 등이 있지만, 일반적으로 구조가 간단하고 안정한 NLMS 알고리즘이 많이 사용된다 [3][4].

음향 반향제거기는 적응 필터의 안정성을 위해 마이크 입력신호에서 사용자의 음성신호를 검출하는 동시통화 검출기가 필요하며, 검출결과로부터 적응필터 탭 계수의 갱신 여부를 판단한다. 보다 정확한 DT 검출은 음향 반향제거기의 성능에 크게 영향을 미치기 때문에 DT 검출에 사용되는 알고리즘 또한 여러 방향으로 연구되어왔다. 즉, 원단화자와 마이크 입력신호의 전력을 비교하는 방법 [5], 선형예측계수 (LPC : Linear Prediction Coefficient)를 이용하는 방법 [6], 그리고 상호상관계수를 이용한 방법 [7] 등이 있다. 그중 상호상관계수를 이용한 방법은 연산량이 적으면서 성능이 우수하여 DT 검출에 많이 이용된다. 상관계수를 이용한 동시통화기간 검출을 비롯한 여러 DT 검출 알고리즘은 반향 경로가 바뀐 EPC (Echo Path Change) 구간이나 배경잡음이 존재하는 구간에서 DT 검출 오류가 상당량 발생한다. 이는 EPC의 경우 순간적으로 변한 반향경로로 인해 오차신호에 제거하지 못한 반향신호가 존재하게 되기 때문이며, 잡음환경 하에서 기준신호로부터 추정하지 못하는 배경잡음을 DT 신호로 검출하기 때문이다. 이를 극복하기 위해 반향

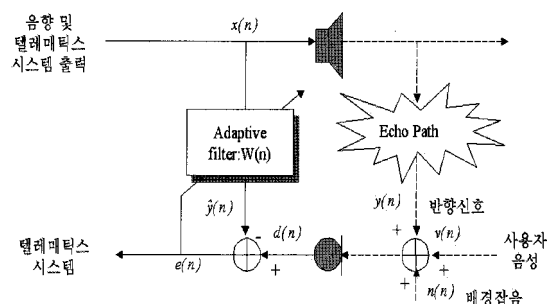


그림 1. 적응 음향 반향제거기의 구조  
Figure 1. The structure of an adaptive echo canceller.

경로를 추정하는 적응필터 외에 보조필터를 사용하는 방법 [8]이 제시되기도 하였으며, 마이크로폰 배열 (array) 을 이용한 반향제거 시스템이나, 잡음제거 기법 등이 소개되기도 하였으나 [9][10], 만족할 만한 성능을 얻지 못하였다.

### III. 제안한 음성인터페이스

본 논문에서는 반향제거기와 더불어 사용자의 음성 유무에 따라 출력신호를 제어하는 barge-in 기능을 갖는 차량용 음성인터페이스를 제안한다. 그림 2는 제안한 음성인터페이스의 전체적인 시스템 블록 다이어그램이다. 배경잡음에 강인한 DT 검출 성능을 위해 앞서 소개한 보조필터와 함께 배경잡음과 반향신호의 평균 에너지 파라미터를 제시한다. 더불어 시간에 따라 변하는 배경잡음과 반향신호의 영향을 고려하여 적응필터 계수의 갱신 구간을 설정하였으며, 배경잡음 및 잔여 반향성분 제거를 위한 후 처리단으로 MMSE-STSA (Minimum Mean Square Error Short Time Spectral Amplitude) 기법 [11] 과 가변의 임계치를 가지는 센터클리핑 (center clipping) 을 적용하였다.

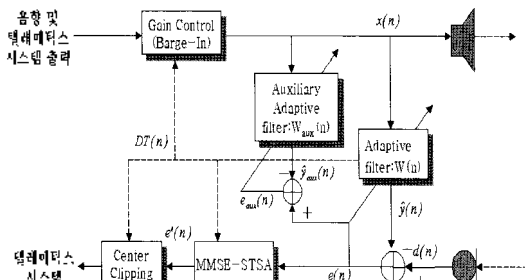


그림 2. 제안한 음성인터페이스 시스템의 블록 다이어그램  
Figure 2. Block diagram of proposed speech interfac.

#### 3.1. DT 검출 알고리즘

상호상관계수를 기준으로 DT를 검출하는 반향제거기의 경우, 배경잡음은 기준신호로부터 추정할 수 없는 신호로서 입력신호와 오차신호와의 상관계수가 많은 곳에서 높게 나타나며, 이는 사용자의 음성입력과 더불어 DT 신호로 검출된다. 하지만, 배경잡음이 우세한 부분과 사용자 음성이 우세한 부분은 그림 3과 같이 입력신호의 에너지를 비교해 봄으로써 구분할 수 있다. 즉, 그림 3(c) 및 3(d)에서 보듯이 사용자 음성명령에 의한 DT 구간 (실선)은 상관계수 값과 입력신호의 평균에너지가 모두 크

게 나타나지만, 배경잡음이 반향신호보다 크거나 유사한 경우 (점선) 상관계수는 크지만 입력신호의 평균 에너지 값은 다소 낮은 수치를 가진다. 적절한 문턱값을 설정하여 두 경우를 구분할 수 있지만, 잡음과 반향신호가 시간에 따라 변하는 값을 가지기 때문에 고정된 문턱값을 정하기가 쉽지 않다.

따라서 본 논문에서는 배경잡음에 의한 DT 검출 오류를 방지하기 위해 입력신호 평균 전력  $\sigma_d^2(n)$ 의 재귀적인 평균값에 해당하는 새로운 파라미터  $Pow_{TH}$ 를 정의하고, 이를 이용하여 DT 검출을 재확인하는 과정을 수행하였다.  $Pow_{TH}$  파라미터는 마이크로폰으로 입력되는 신호의 평균전력  $\sigma_d^2(n)$ 을 이용하여 식 (1)~(3)과 같이 정의한다.

$$\sigma_d^2(n) = (1 - \beta_1) \cdot \sigma_d^2(n-1) + \beta_1 \cdot d^2(n), \quad \beta_1 = 1/256 \quad (1)$$

$$\sigma_d''(n) = (1 - \beta_2) \cdot \sigma_d''(n-1) + \beta_2 \cdot \sigma_d^2(n), \quad \beta_2 = 1/256 \quad (2)$$

$$Pow_{TH}(n) = (1 - \beta_3) \cdot Pow_{TH}(n-1) + \beta_3 \cdot \sigma_d''(n), \quad \beta_3 = 1/128 \quad (3)$$

단, 사용자의 음성에 의한 입력신호의 평균전력 변화를 무시하고 배경잡음과 반향신호의 평균전력만을 고려하기 위해 DT 구간에서는  $Pow_{TH}$  파라미터를 갱신하지 않는다. 또한 각 전력 파라미터의 계산에 있어 서로 다른 망각지수  $\beta$ 를 적용하여  $Pow_{TH}$  파라미터가 배경잡음과 반향신호에 의한 평균전력의 영향을 최대한 고려하도록 하였다.

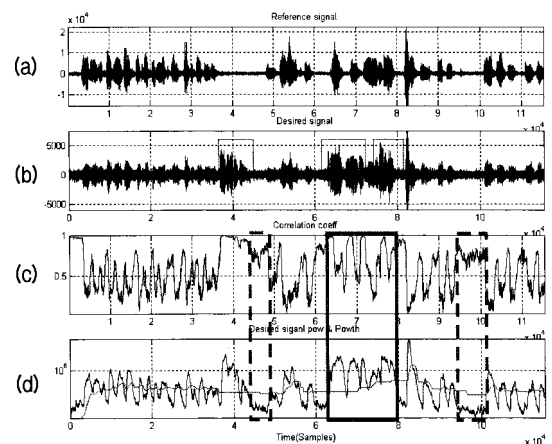


그림 3. 배경잡음에 따른 상호상관계수와 평균에너지 : (a) 기준 신호 (b) 입력신호 및 DT 검출 결과 (ESNR 5dB) (c) 상호상관계수 값 (d) 입력신호의 전력과  $Pow_{TH}$

Figure 3. Correlation value and average energy in noisy environment : (a) Reference signal (b) Desired signal and DT detection (ESNR 5 dB) (c) Correlation value (d) Input signal power and  $Pow_{TH}$

그림 4와 5는 본 논문에서 제안한 DT 시작점 및 끝점 검출 알고리즘의 순서도이다. DT의 시작점 검출을 위해 입력신호와 오차신호의 상호상관계수의 값을 임계치 ( $TH = 0.5$ )와 비교하고, 추가적으로 입력신호의 평균 에너지로부터 추정된 에너지 파라미터와 비교함으로써 잡음 환경에서 보다 정확히 DT를 검출할 수 있다. 더욱이 검출된 DT 구간에서 보조필터를 동작시켜 EPC에 의한 DT 검출 오류를 재확인함으로써 성능을 향상시켰다. DT로 검출되지 않은 구간은 적응 필터가 반향경로를 추정하는 구간 (CNVG : Convergence)으로 추정된 반향 경로에 의해 오차신호가 수렴하게 된다.

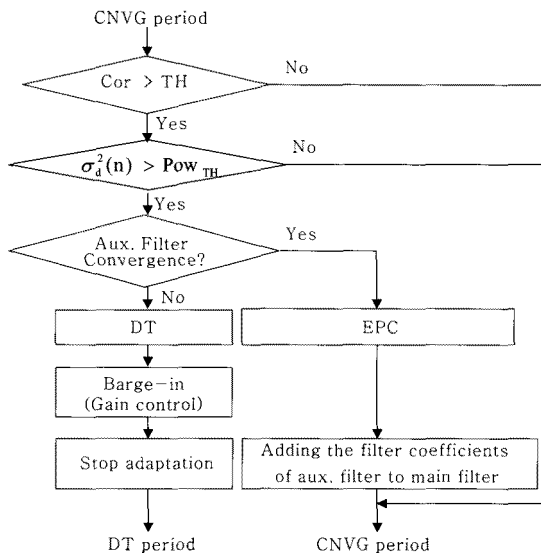


그림 4. 제안한 DT 시작점 검출 알고리즘  
Figure 4. Proposed DT start-point detection algorithm.

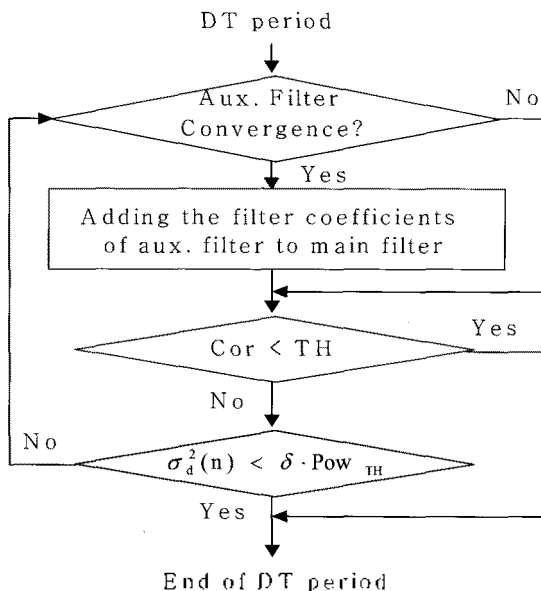


그림 5. 제안한 DT 끝점 검출 알고리즘  
Figure 5. Proposed DT end-point detection algorithm.

DT의 끝점 검출은 그림 5와 같이 크게 3단계로 이루어진다. 우선적으로 EPC에 의한 DT 검출 오류를 찾아내기 위해 보조필터의 수렴여부를 확인하고, 실제 DT의 끝점 검출 방법으로 상호상관계수 값을 확인한다. 더불어 마이크로폰의 입력신호의 평균 에너지 값을 추정된 에너지 파라미터에 실험적으로 구한 가중치 ( $\delta = 1.5$ )를 곱한 값과 비교함으로써, 입력신호가 사용자의 음성신호를 포함하지 않는 지점, 즉 입력신호의 에너지가 추정된 에너지 임계치보다 낮은 DT의 끝점을 검출한다.

### 3.2. 적응필터 갱신 구간 설정

일반적인 반향제거기는 DT 구간을 제외한 모든 구간에서 적응필터의 계수를 갱신하게 된다. 하지만 반향신호가 배경잡음에 묻혀 버리는 경우, 기준신호로부터 반향신호가 아닌 배경잡음 신호를 추정하게 되며, 이로 인해 적응 필터는 불안정해진다. 따라서 반향신호가 존재하지 않는 영역 및 반향신호가 배경잡음 보다 낮은 영역을 검출하여 적응필터의 갱신을 제한하는 것이 필요하다. 간단한 방법으로 식 (4)와 같이 출력되는 기준신호의 평균 에너지 값  $\sigma_r^2(n)$ 을 설정한 임계치와 비교함으로써 판단할 수 있다. 하지만 고정된 임계치를 사용할 경우, 실제 존재하는 배경잡음의 크기와는 상관없이 적응필터 갱신 구간이 결정되기 때문에 배경잡음이 큰 구간에서는 적응필터의 오추정을 막지 못하는 반면, 배경잡음이 작은 영역에서는 반향 경로의 추정 시 사용할 수 있는 구간에서도 적응필터 계수를 갱신하지 못하는 경우가 발생한다.

$$\sigma_r^2(n) < 1000 \tag{4}$$

$$\sigma_w^2(n) = (1 - \beta) \cdot \sigma_w^2(n-1) + \beta \cdot e^2(n) \tag{5}$$

$$\sigma_r^2(n) > \gamma \cdot \sigma_w^2(n) \tag{6}$$

따라서 반향신호를 제거한 오차신호로부터 식 (5)와 같이 배경잡음의 평균 에너지를 계산하고 기준신호의 에너지와 비교함으로써 식 (6)과 같이 기준신호의 평균에너지가 추정된 배경잡음 에너지보다 큰 구간을 적응필터 갱신 구간으로 설정하였다. 식 (5)에서  $\beta$ 는 망각지수로서 한 프레임 길이의 역수 값 0.0039를 나타내며, 식 (6)의  $\gamma$ 는 시스템 출력장치로부터 마이크 입력까지 반향경로에서 발생하는 감쇠의 역수에 해당하는 파라미터로 추정된 반향경로에 대한 임펄스 응답으로부터 값을 추정할 수 있다.

### 3.3. Barge-in 기능

안내 메시지의 출력 종료 여부를 떠나 사용자가 명령을 입력할 수 있도록 사용자의 음성신호가 검출되면 음향 및 텔레메틱스 시스템의 출력신호를 줄이거나 꺼버리는 barge-in 기능은 정확한 DT 검출이 이루어진다면 그림 6과 같이 시스템의 출력 이득을 제어함으로써 간단히 구현할 수 있다 [12].

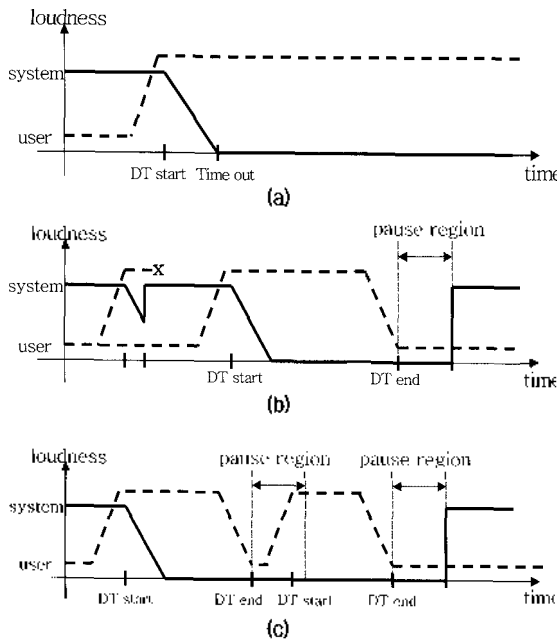


그림 6. Barge-in의 이득제어 : (a) 음성 신호 입력 시 (b) DT 검출 오류와 DT 검출 시 (c) 연속된 DT 구간

Figure 6. Gain control of barge-in functionality : (a) Speech signal input (b) DT detection and DT detection error (c) Continuous DT region.

## IV. 실험 및 결과고찰

### 4.1. 실험 환경 및 조건

실제 환경에서 잡음 및 반향신호의 크기에 따른 다양한 환경을 구현하여 실험하기가 쉽지 않기 때문에 가상의 차량 실내 환경을 만들고 image method [13]를 이용한 룬 임펄스 응답을 적용하여 원하는 신호를 만들어 시뮬레이션을 수행하였다. 모의실험의 환경으로 그림 7과 같이 가로×세로×높이를 각각 1.5×2.0×1 m로 자동차의 실내 환경을 모델링 하였으며, 운전자의 입 위치 및 여러 개의 실내 스피커의 위치를 가정하였다. 차량 실내에서의 50 msec 잔향 시간을 가정하였으며, 마이크로폰은 등방성 빔패턴을 가지고 있으며 한쪽 면에 붙어 있어 후면에서 반사되는 반사파는 없다고 가정하였다. 음향 반향

제거기에 사용되는 적응필터 및 보조필터의 필터 길이는 각각 256, 128을 사용하였다 [8].

여러 종류의 시스템 출력 신호에 대한 성능 평가를 수행하기 위해 시스템 출력 신호는 음성 (English, Korean) 과 음악 (한국 팝 2곡) 그리고 연주곡 신호로서 Vivaldi의 사계 중 가을 전반부 음을 사용하여 다양한 조건에서 시스템 출력신호를 생성하였다. 배경잡음은 ETSI (European Telecommunication Standard Institute)의 Aurora2 DB에 포함되어있는 차량용 잡음신호를 선택하여 적용하였으며, 운전자의 발성음성은 간단한 음성 명령이나 음성 통화 등을 가정한 것으로, 간단한 명령어 및 4연 숫자음, 대화체의 문장을 각각 설정하였고, 운전자의 음성신호 존재 정보를 이용하여 시뮬레이션에 의한 DT 검출 결과와 비교 할 수 있도록 하였다.

시뮬레이션의 잡음 환경을 설정하기 위해 ESNR (Echo Signal to Noise Ratio)과 SER (Speech to Echo Ratio)을 식 (8) 및 (9)와 같이 정의하였다. ESNR은 배경잡음에 대한 반향신호의 크기를 나타내며, SER은 반향신호에 대한 음성신호의 크기를 결정한다. 이로서 다양한 반향 및 잡음 환경을 정의하고 실험을 통해 배경잡음의 정도에 따른 DT 검출 및 적응필터의 반향경로 추정에 끼치는 영향을 분석하였다.

$$ESNR[dB] = 10 \cdot \log_{10} \frac{\sum_{n=0}^M y(n)^2}{\sum_{n=0}^M n(n)^2} \quad (8)$$

where  $M$ : signal length

$$SER[dB] = 10 \cdot \log_{10} \frac{\sum_{n=0}^M x(n)^2}{\sum_{n=0}^M y(n)^2} \quad (9)$$

where  $M$ : signal length

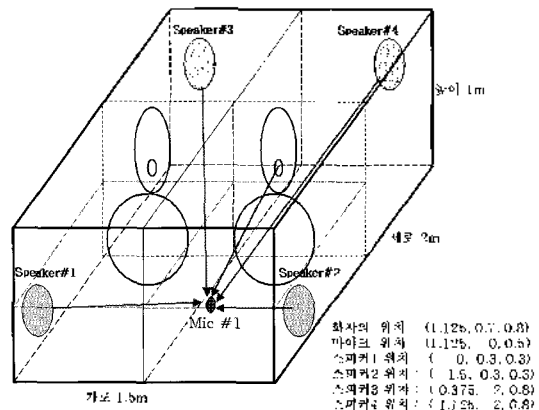


그림 7. 실험을 위한 시뮬레이션 환경  
Figure 7. Simulation environment for experiment.

제한한 음성인터페이스의 성능 평가를 위해 DTFR (Double Talk False-detect Rate), ERLE (Echo Return Loss Enhancement)의 평균값을 성능 분석 파라미터로 정의하고 비교하였다. DTFR는 실제 DT 검출 오류가 얼마나 많이 발생하는가를 판단하는 파라미터로서, 알고 있는 운전자의 신호로부터 정확한 DT 정보  $DT_{original}(n)$ 을 추출하고, 알고리즘을 통해 검출된 DT 결과인  $DT_{detect}(n)$ 에 대해 DT 검출 오류를 식 (10)과 같이 배타적 논리합 연산 ( $\oplus$ )을 이용하여 계산한다. 계산된 검출오류를 식 (11)과 같이 전체 신호에 대한 검출 오류 비율을 나타내는 DTFR로 분석할 수 있다. 정확한 DT 검출이 이루어질수록 DTFR 값은 낮아 질 것이다. ERLE는 반향제거기의 성능 평가를 위해 사용되는 가장 잘 알려진 파라미터로서 식 (12)와 같이 정의되며, 시스템의 전반적인 반향제거 성능 분석을 위해 각 신호에 대한 ERLE의 평균값을 잡음 환경에서 성능비교 파라미터로 사용하였다.

$$DT_{error}(n) = (DT_{detect}(n) \oplus DT_{original}(n)) \quad (10)$$

$$DTFR = \left( \frac{1}{N} \cdot \sum_{n=0}^{N-1} DT_{error}(n) \right) \times 10 \quad (11)$$

$$ERLE(n) = 10 \cdot \log_{10} \left( \frac{\sum_{i=0}^{L-1} |d(n-i)|^2}{\sum_{i=0}^{L-1} |e(n-i)|^2} \right) \quad (12)$$

where  $L$ : frame length

### 4.2. 반향 제거 및 DT 검출 실험

제한한 반향 제거기의 성능 평가를 위해 여러 배경잡음의 크기에 따른 시뮬레이션 결과를 아래의 기존 2가지 알고리즘과 비교하였다.

- 알고리즘 1 : 상호 상관도 (Correlation)를 이용한 DT 검출 및 고정된 문턱값을 이용한 적응필터 갱신 구간 설정
- 알고리즘 2 : 보조필터를 이용한 DT 검출 알고리즘 및 고정된 문턱값을 이용한 적응 필터 갱신 구간 설정
- 알고리즘 3 : 보조필터와 전력 파라미터  $P_{ow,TH}$ 를 이용한 DT 검출 및 오차신호에서의 배경잡음 전력  $\sigma_w^2(n)$ 를 이용한 적응 필터 갱신 구간 설정 (제안한 알고리즘)

실험에서는 여러 배경잡음 환경을 위해 ESNR을 5 dB ~ 30 dB (5 dB 간격) 범위에서 변화시켰으며, SER은 5dB로 설정하고 시뮬레이션을 수행하였다. 각각의 반향제거 방식에 대해 각각의 시스템 출력신호 및 잡음 환경에서

동일하게 실험하였으며 그 결과는 그림 8 및 9와 같다.

제시한 결과 그래프에서도 볼 수 있듯이 제안한 알고리즘에서 DT 검출 오류가 크게 줄어들었음을 확인할 수 있다. 특히 ESNR 수치가 낮은 경우에도 다른 기법들에 비해 DT를 보다 정확히 찾는 것을 확인할 수 있다. 일반적으로 ERLE 특성은 NLMS 적용 알고리즘의 특성에 많은 영

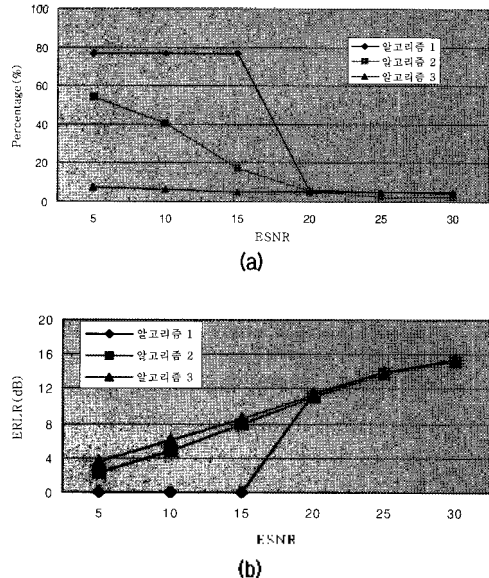


그림 8. 시스템 출력신호 1에 대한 실험 결과 : (a) ERLE (b) DTFR  
Figure 8. Experiment result for system output signal 1 (a) ERLE (b) DTFR.

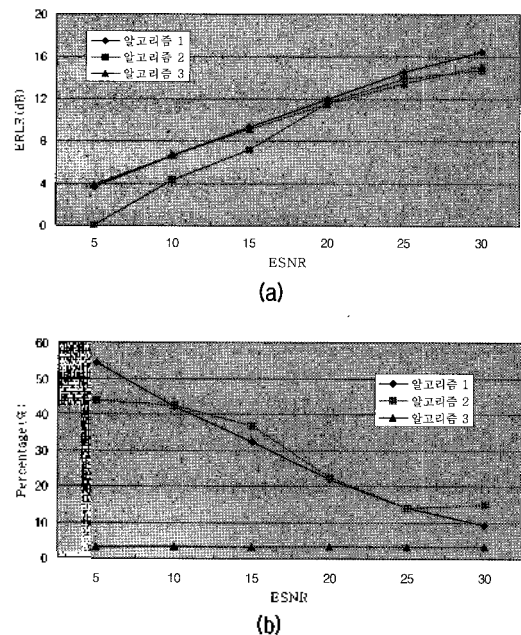


그림 9. 시스템 출력신호 2에 대한 실험 결과 : (a) ERLE (b) DTFR  
Figure 9. Experiment result for system output signal 2 (a) ERLE (b) DTFR.

향을 받기 때문에 각 방법에 대해서 그 값의 변화가 크지 않다. 하지만, 제안한 알고리즘의 경우 배경잡음의 크기에 따라 적응필터의 갱신구간을 설정함으로써 적응 필터의 안정성이 향상되기 때문에 ESNR이 낮은 영역에서 다른 기법들에 비해 다소 높은 ERLRE를 가지는 것을 확인할 수 있다.

### 4.3. 실제 연구실 환경에서의 실험

제안한 음성인터페이스 알고리즘을 그림 10과 같이 연구실 환경에서 구현하고 성능을 확인하였다. PC를 음향 시스템으로 간주하고 스피커로 출력되는 신호에서 기준

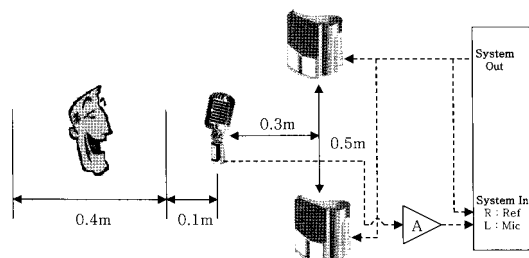


그림 10. 음성인터페이스 실험 환경  
Figure 10. Experiment environment of speech interface.

신호를 구하고, 공간상의 반향경로를 거친 마이크 입력 신호는 기준신호와와의 신호 레벨을 맞추기 위해 증폭기를 사용하여 증폭한 신호를 시스템의 입력신호로 PC의 라인 인 단자를 통해 좌/우 두 채널로 입력한다.

그림 11은 제안한 시스템을 실제 연구실 환경에서 실험한 결과이다. 시스템에서 출력되는 신호 (a)는 공간상의 반향경로를 통해 마이크로폰으로 입력된다. 입력된 신호 (b)는 출력신호를 기준신호로 한 적응필터를 거쳐 반향성분이 제거된 신호 (c)를 만들게 된다. 만약 입력신호가 (b)에서 처럼 사용자의 음성이 존재 할 경우, 시스템의 DT 검출기는 신호 (c)와 (f)에서 나타낸 것과 같이 상호상관계수 (Cor)와 추정된 임계치 에너지 ( $Pow_{TH}$ )를 통해 DT 상황을 검출하게 되고 (a)의 출력신호에 barge-in 기능을 적용하게 된다. 최종적으로 반향이 제거되고 후처리 단을 거친 신호는 그림 11의 (c)처럼 사용자의 음성신호만이 남게 된다. 앞서 제시한 반향제거 및 DT 검출 알고리즘으로 DT 시작점 및 끝점 검출을 보다 정확히 수행함을 확인할 수 있으며, (e)의 ERLRE를 통해 반향제거 성능을 확인할 수 있다.

## V. 결론

본 논문에서는 차량용 텔레메틱스 시스템에서 적용 가능한 barge-in 기능과 음향 반향제거기를 갖는 음성인터페이스의 성능 향상 기법을 제안하고 PC 환경에서 구현하였다. 일반적인 음향 반향제거기 및 DT 검출기는 차량과 같이 배경잡음이 존재하는 환경에서 그 성능을 최대로 발휘하지 못한다. 하지만 구현된 시스템은 배경잡음 및 반향 에너지 값의 평균값을 이용하여 DT 검출의 비교 에너지 파라미터로 사용하여 기존의 상호상관계수에만 의존한 DT 시작점 및 끝점 검출 성능을 향상시켰으며, 이를 바탕으로 barge-in 기능을 적용함으로써 시스템은 보다 깨끗한 운전자의 음성을 얻을 수 있다. 더욱이 시스템 출력신호와 배경잡음을 비교하여 적응필터 탭의 갱신 여부를 결정함으로써 적응필터의 발산을 사전에 방지하였다. 구현한 시스템의 성능 평가를 위해 ESNR 및 SFR로 정의되는 배경 환경 변수를 이용하여 가상의 시뮬레이션을 통해 여러 배경잡음 환경에서 시스템 성능을 평가하였다. DTFR과 평균 ERLRE 수치로서 DT 검출 오류 비율과 반향 제거 성능을 분석한 결과, ESNR이 15dB 이하로 배경잡음을 많이 포함한 경우 DTFR 수치가 평균 10% 이하로 나타남을 확인하였으며, 더불어 평균 ERLRE는 5~10

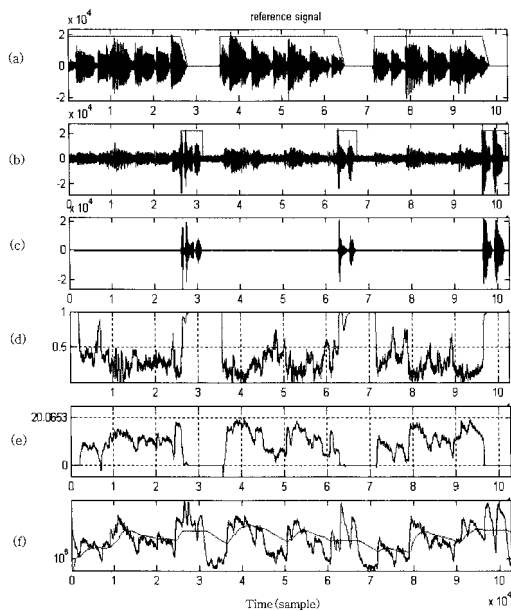


그림 11. 구현 실험 결과 : (a) 기준신호와 시스템 출력이득 (b) 시스템 입력신호와 DT 검출 결과 (c) 처리된 결과 신호 (d) 상호상관계수 값 (e) ERLRE (f) 입력신호의 전력과  $Pow_{TH}$

Figure 11. Implementation experimental result : (a) Reference signal and system output gain (b) Desired signal and DT detection result (c) Processed result signal (d) Correlation value (e) ERLRE (f) Input signal power and  $Pow_{TH}$ .

dB 정도의 이득이 있음을 확인할 수 있었다. 본 논문에서는 구현한 시스템의 성능을 실제 연구실 환경에서의 실험을 통해서도 기존의 반향 제거기에 비해 DT 검출 성능이 뛰어나고 반향신호 또한 잘 제거됨을 확인하였다.

### 참고 문헌

1. Y. Zhao, "Telematics: safe and fun driving," *Intelligent Systems, IEEE Expert*, vol. 17, no. 1, pp. 10-14, 2002.
2. VOICE RECOGNITION tips for OnStar Personal Calling, [http://www.onstar.com/us\\_english/downloadable/opc\\_voice\\_recognition.pdf](http://www.onstar.com/us_english/downloadable/opc_voice_recognition.pdf).
3. S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, New Jersey, 1996.
4. B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, New Jersey, 1985.
5. D. G. Messerschmit, "Echo Cancellation in Speech and Data Transmission", *Selected Areas in Communications, IEEE Journal on*, vol. 2, no. 2, pp. 283-297, 1984.
6. Jae Ha Yoo and Sung Ho Cho, "A New Double Talk Detector Using The Lattice Predictors For An Acoustic Echo Canceller", in *Proc. IEEE, Conf. TENCON'97, Speech and Image Technologies for Computing and Telecommunications*, vol. 2, pp. 483-486, Dec. 1997.
7. Hua Ye and Bo-Xiu Wu, "A New Double-Talk Detection Algorithm Based on the Orthogonality Theorem", *IEEE Trans, Communications*, vol. 39, no. 11, pp. 1542-1545, 1991.
8. 김시호, 권홍식, 배건성, "음향반향제거기에서 보조필터를 이용한 동시통화 검출 성능 개선", *한국음향학회지*, 21권, 1호, 249-252쪽, 2002.
9. M. Kallinger and J. Bitzer, "Study on Combining Multi-Channel Echo Cancellers with Beamformers", in *Proc. Int. Conf. Acoust. Speech and Signal Processing, ICASSP'00*, vol. 2, pp. 797-800, June, 2000.
10. H. Fuder and P. Dreiseitel, "Implementation of A Hands-Free Car Phone with Echo Cancellation and Noise Dependent Loss Control", in *Proc. Int. Conf. Acoust. Speech, and Signal Processing, ICASSP'00*, vol. 6, pp. 3622-3625, June, 2000.
11. Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator," *IEEE Trans, Acoust. Speech And Signal Processing*, vol. 32, no. 6, pp. 1109-1121, 1984.
12. N. Ström and S. Seneff, "Intelligent barge-in in conversational systems" *MIT laboratory for Computer Science, Proc. ICSLP*, Beijing, China, Oct. 2000.
13. Jont B. Allen and David A. Berkley, "Image method for efficiently simulating small-room acoustics", *Journal of ASA*, Vol. 65, no. 4, pp. 943 - 950, 1979.

### 저자 약력

#### •김 준 (Jun Kim)



2003년 2월: 경북대학교 전자전기컴퓨터학부 (공학사)  
 2005년 2월: 경북대학교 전자전기컴퓨터학부 (공학 석사)  
 2005년 2월~ 현재: 국방과학연구소 무인지능화연구실

#### •배 건 성 (Keunsung Bae)



1977년 2월: 서울대학교 전자공학과 (공학사)  
 1979년 2월: 한국과학기술원 전기및전자공학과 (공학 석사)  
 1989년 5월: University of Florida (공학박사)  
 1979년 3월~ 현재: 경북대학교 전자전기컴퓨터학부 교수