

무전기 음성통신에서 최적음성채널 선택을 위한 개선방안에 관한 연구

류창국* · 이배호**

Study on Improvement for selecting the optimum voice channels in the radio voice communication

Chang-Guk Lew* · Bae-Ho Lee**

요 약

지상 관제소에 근무하는 항공관제사와 비행중인 항공기는 무전기를 이용하여 음성통신을 한다. 항공기에서 송신하는 음성신호는 전국에 있는 다수의 지상사이트에 동시에 수신된다. 이때 항공관제사는 항공기와의 거리, 속도, 기상상태, 안테나와 무전기 조정상태 등에 따라 다양한 품질의 음성신호를 수신하게 된다. 항공관제사는 매 순간 최적의 음성신호를 찾아 항공기와 최적의 상황에서 음성통신을 수행한다. 그러나, 현재는 입력된 음성의 음량(Gain)을 기준으로 CD(Carrier Dectect)값이 우수하다고 판단되는 신호를 최적채널로 선택하지만, 이는 잡음이 통화품질에 미치는 영향을 고려하지 않기에 최적채널을 선택한다고 볼 수 없다. 본 논문을 통해 수신된 음성신호에서 잡음을 제거한 후 사용자가 최적채널을 선택할 수 있도록 수치화된 정보 및 개선된 음질의 음성신호를 제공할 수 있었다. 이를 이용하여 항공기 관제 또는 훈련감청시스템 운용 시 향상된 품질의 채널을 선택하여 안전사고 예방, 훈련 능력향상 등을 기대할 수 있다.

ABSTRACT

An aircraft in flight and ATC(Air Traffic Controllers) working in the Ground Control Center carry out a voice communication using the radio. Voice signal to be transmitted from the aircraft is received to a plurality of terrestrial sites around the country at the same time. The ATC receives the various quality of voice signal from the aircraft depending on the distance, speed, weather conditions and adjusted condition of the antenna and the radio. The ATC carries out a voice communication with aircraft in the optimal conditions finding the best voice signal. However, the present system chooses the values of the CD(Carrier Dectect) which is determined to be superior to, based on the input voice level, as optimal channel. Thus this system can not be seen to select the optimal channel because it doesn't consider the effect of the noise which influences on the communication quality. In this paper, after removing the noise in the voice signal, we could give the digitized information and an improved voice signal quality, so that users can select an optimal channel. By using it, when operating the training eavesdropping system or the aircraft control, we can expect prevention accident and improvement of training performance by selecting the improved quality channel.

키워드

MOS, Air Traffic Control, RADIO Communication, Military Communication

음성 품질, 항공 관제, 무선 통신, 군통신

* 전남대학교 전자컴퓨터공학(ds1gor@naver.com)

** 교신저자 : 전남대학교 전자컴퓨터공학부

• 접수일 : 2015. 12. 05

• 수정완료일 : 2016. 02. 13

• 게재확정일 : 2016. 02. 24

• Received : Dec. 05, 2015, Revised : Feb. 13, 2016, Accepted : Feb. 24, 2016

• Corresponding Author : Bae-Ho Lee

Dept. Chonnam National University, Electronics & Computer Engineering

Email : bhlee@chonnam.ac.kr

I. 서론

국내 항공관제는 VHF, UHF 대역을 이용하여 비행 중인 항공기와 전국의 사이트 간에 무선 음성통신을 수행하고 이를 유선으로 관제소에 전달한다. 그리고, 육군 과학화전투훈련단에서 훈련 시 이동 무전국 간의 통신내용을 주변의 사설 기지국에서 감청한다. 두 상황 모두 무전기를 통해 대기를 전달매개로 하여 신호가 전달되기 때문에 시시각각 전파상황과 안테나와 무전기의 조정상태에 따라 다양한 품질의 음성신호가 상호간에 전달된다. 따라서 항공관제사와 훈련상황 감청자는 최고 품질의 음성신호를 선택하여 청취하고자 하는데, 현재는 수신되는 음성의 음량으로만 품질을 평가하여 선택한다. 이때 수신되는 음성은 송신하고자 하는 음성신호, 주변잡음, 통신잡음의 합이다. 따라서, 단순히 음량의 크기는 최적의 신호로 처리하기엔 여러 문제점이 있다. 음량의 크기가 아닌 수신되는 음성신호의 음질을 평가하여 이를 사용자와 시스템에 전달할 필요가 있다.

음질 평가는 주로 청취자들에 의해서 직접 평가되는 주관적 음질 평가방법과 수학적 계산으로 얻어지는 객관적 음질 평가법이 있는데, 이중 주관적 음질 평가방법은 사람에 의해서 직접적으로 평가되므로 실제 사용자의 체감 음질을 가장 잘 표현할 수 있다. 그러나 주관적 음질 평가는 시간과 비용이 많이 소모되므로 객관적 음질 평가 척도를 이용하여 주관적 음질을 예측하는 것이 보다 바람직하다[1]. 본 논문에서는 입력된 음성신호를 기반으로 신호처리 전후의 신호를 비교 후 음질을 판별하여 MOS와 같이 등급을 나누고[2], 제시된 품질값을 기준으로 사용자가 판별하여 선택할 수 있는 방안을 제시하고자 한다.

II. 관련연구

음성 품질평가기법[2]은 대표적으로 주관적 평가방법과 측정데이터로부터 객관적인 자료를 이용해 성능을 평가하는 방법이 있다. 주관적인 평가방법은 사람

이 직접 청취한 후에 품질을 평가하는 방법이기 때문에 평가자의 내, 외부적인 영향에 의해 평가기준이 달라질 수 있어 신뢰성 확보가 어렵다. 객관적 평가방법은 신호의 입력과 출력에 대한 측정자료나 계산식을 기반으로 평가하기에 주관적인 방법보다는 신뢰성 측면에서는 유리하다.

음성통신 품질은 전화 등을 이용한 통화에서 송화자의 입과 수화자의 귀 사이에 전달되어지는 신호의 품질의 정도를 의미한다. 통화의 명료성에 대한 척도의 기준이 되는 요소로는 수화음량(Loudness), 명료도(Articulation), 반복도(Repetition), 이해도(Intelligibility), Opinion Test 등이 있다. 또한, 통화 품질에 영향을 주는 주요 요인들로는 아날로그 전송계에서는 누화, 반향, 군지연왜곡, 감쇄왜곡 등이 있고, 디지털 전송계에서는 지연, 지연지터, 패킷손실 등이 있다. 이런 원인으로 수신되는 음성은 송신되는 음성에 비해 점차로 악화와 왜곡되어 전달된다.

항공관제와 훈련감청 시스템에서는 매 순간 음질을 평가하여 최적의 채널을 선택하여 음성통신을 해야 하는데, 음성신호의 품질의 평가가 아닌 음량으로 판단하고 선택하고 있어, 잡음이 큰 음량으로 입력되는 경우도 우수한 채널로 선택되는 경우가 종종 있다.

2.1 대표적인 음성 품질평가기술

MOS(Mean Opinion Score, ITU-T P.800)는 음성 코덱에 의해 음성신호를 압축 과정과 네트워크 전송 후에 수신된 음성 품질에 대해서 1점에서 5점까지 5단계로 나누어서 미리 선정된 청취자 집단에 의해 주관적인 점수로 결정된다[3].

PESQ(Perceptual Evaluation of Speech Quality, ITU-T P.862)는 음성 품질에 대한 객관적인 평가 방법으로써 원래의 신호와 원래 신호가 통신 시스템을 통과하여 나온 감쇄된 신호를 평가한다[3].

PEAQ(Perceptual Evaluation of Audio Quality, ITU-T BS.1387)는 두 개의 신호를 비교하여 청감상 느끼게 되는 에러를 산출하는 알고리즘이다[4].

E-Model(ITU-T G.107)은 ESTI TC STQ(Technical Committee for Speech Processing, Transmission & Quality Aspects)은 망을 설계하고자 하는 사람이 실제 망을 설치하기 전에 망의 품질을 계산식을 통해 추정하기 위한 설계도구로 개발하였다[5].

1) Mean opinion score (MOS)

https://en.wikipedia.org/wiki/Mean_opinion_score

2.2 항공관제 시스템의 문제점

그림1은 현재 항공관제사가 운용하는 화면의 일부로 'Assign'된 채널로 부터 음성이 수신되면 CD가 점등하며, 수신레벨을 가장 열악한 '1'에서 가장 우수한 신호인 '5'까지 표시하도록 되어 있다. 이때 수신레벨은 관제사에게 필요한 음성신호뿐만 아니라, 각종 잡음이 포함된 신호로 수신레벨이 '5'라고 해도, SNR이 나쁜 신호가 수신될 경우도 있어, 해당 채널이 최적의 채널이 아닐 수도 있다. 심지어는 음성신호 없이 잡음만 수신되는 경우도 있다.



그림 1. 항공관제 화면 예

Fig. 1 Example of air traffic control screen

2.3 훈련감청 시스템의 문제점

육군 과학화전투훈련단에서는 정보, 통신, 광학, 컴퓨터 시뮬레이션 등의 첨단과학기술을 통합하여 대대급 2개의 팀으로 구성하여 모의전투를 한다. 실제 전장과 유사한 훈련환경을 구축하고 훈련통제본부에서는 이들 2개 팀의 전투자료인 음성과 영상을 실시간으로 수집하고, 이를 분석 및 사후검토의 과정을 통해 훈련효과를 극대화하고 있다²⁾.

이때, 대대급 훈련에서 발생하는 무선통신을 주변의 기지국에서 감청해서 최적채널을 선택해서 훈련통제본부에서 감청하도록 운영한다. 그림2과 같이 입력된 음성을 ADC(Analog to Digital Converter)하여 음성신호의 크기로 수신감도의 레벨을 최적채널로 선택하여 사용자에게 제공하였으나, 음성신호뿐만 아니라 각종 잡음과 불필요한 신호가 혼합된 신호를 선택할 가능성이 있기에 신호의 레벨은 낮아도 SNR이 좋은 신호가 선택되지 않는 문제 있다.

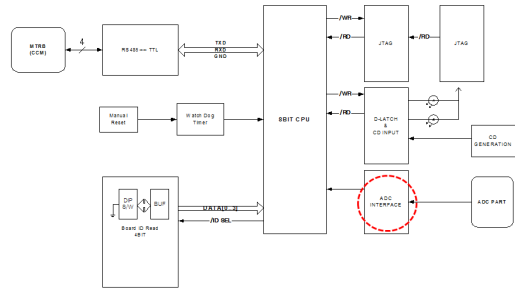


그림 2. 감청장비 채널선택 예

Fig. 2 Example of eavesdropping equipment channel selection

III. 제안한 개선방법

3.1 음성 신호처리를 이용한 음질평가

음성의 품질을 좌우하는 다양한 형태의 잡음의 유입은 VHF, UHF의 무선통신 환경의 특성에 기인할 수도 있지만, 통화자의 다양한 환경에서 발생하는 주변잡음이 마이크에 혼입되어, 송화자의 음성과 믹싱되어 수신자에게 전달되어 전체적인 통화품질을 악화시킨다. 송화자의 환경이 항공기의 조종사인 경우 기체에서 발생하는 엔진잡음, 바람소리, 주변기계음, 주위의 다른 사람의 통화소리, 기체조작 소음 등 다양한 형태의 통화와 관계없는 잡음이 발생하고, 이는 통화 중에 수신자에게 전달되어서 정확한 음성수신을 어렵게 한다.

이러한 잡음이 포함된 신호를 분석하여 잡음을 제거한다. 수신된 다수의 음성신호 중에 최적의 신호를 사용자에게 제시하여 이를 선택할 수 있도록 SNR에 대한 객관적 지표를 제시함으로써, 주관적 판단이 아닌 수치화된 정보를 제공한다. 음성의 품질을 사용자에게 제시함으로써 자동 또는 수동으로 최상의 음성신호를 선택하여 사용할 수 있다.

2) <http://www.kctc.mil.kr/Training1.html>

표 1. 음성신호의 MOS 5단계 평가척도[2]

Table 1. MOS 5 level evaluation scale of the speech signal[2]

MOS	Quality Proposal	MOS Quality	Impairment
5	90~100	Excellent	Imperceptible
4	80~99	Good	Perceptible but not annoying
3	70~89	Fair	Slightly annoying
2	60~79	Poor	Annoying
1	0~59	Bad	Very annoying

일반적으로 MOS 4~5 정도이면 고품질, ISDN호 품질 이상, Toll 품질이라고 하고, MOS 3.5~4 정도를 중간품질, 32kbps ADPCM 품질정도, 자연스러운 통화를 할 수 있는 통화수준의 품질이고, MOS 3~3.5 정도는 대화가 잘 이루어지지만 품질 저하를 느낄 수 있으며, MOS 2.5~3 정도를 군사품질, 대화 가능하지만 집중할 정도의 품질로 규정하고 있다.

3.2. 제안한 알고리즘

음성이 입력되면, 음성레벨을 측정하여 특징을 분석하고, 지역잡음왜곡을 측정 분석한 후, 전역잡음을 측정 분석한 후, 불균형 음성품질을 분석하고, 피치교차파워[6] 분석하여 음성의 특색, 남성과 여성, 고음, 저음 등의 특징을 분석하여 이를 일정한 길이의 프레임으로 만들어서 음질을 평가한다.

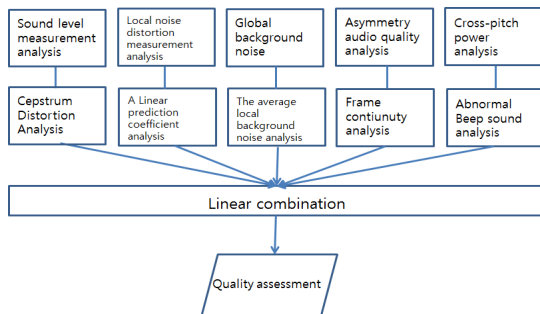


그림 3. 시스템 블럭도

Fig. 3 A system block diagram

이때 전강조필터(Pre-emphasis)를 사용하는데 이는 입력되는 신호를 일정한 크기의 형태로 가공하는 과정이다. 식(1)를 이용하여 AGC(Auto Gain Control)한 후 출력이 일정한 신호레벨이 되도록 한다.

$$H(z) = 1 - \tilde{a}z^{-1}, 0.9 \leq a \leq 1.0$$

$$\tilde{s}(n) = s(n) - \tilde{a}s(n-1) \tag{1}$$

입력된 음성을 식(1)을 이용하여 전강조필터 처리 전후의 3차원 스펙트럼을 보면 그림[4]와 같다.

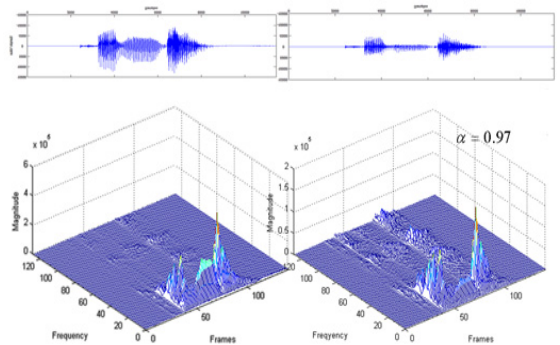


그림 4. 음성신호처리 전후 스펙트럼

Fig. 4 Voice signal processing spectrum before and after

전처리 과정인 전강조필터 처리로 입력된 음성의 에너지 분석과정으로 최적의 음성을 판단하도록 한다.

$$E = \sum_{n=0}^{N-1} x^2(m) \tag{2}$$

$$Z(n) = \frac{1}{2N} \sum_{n=0}^{N-1} [|x(m)| - |x(m-1)|] \tag{3}$$

그림 5의 중간은 시간도메인에서의 입력된 음성신호를 의미하고, 상단은 입력된 음성신호를 식(2)를 이용하여 음성신호의 에너지를 확인하여, 음성신호의 존재를 확인한다.

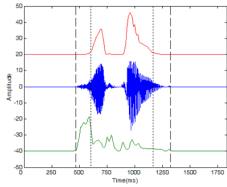


그림 5. 음성신호와 Spectrogram 분석
Fig. 5 Voice signal and spectrogram analysis

이때 불필요한 고주파 성분의 잡음성분을 제거하면, 최적의 음성신호가 추출되고 이를 판단의 기준으로 삼는다. 음성신호의 초기 입력은 복잡한 성분으로 표현되나, 이 신호가 로우패스 필터를 통과하면 보다 단순한 형태의 음성신호로 정확한 판단을 할 수 있다. 이는 음성의 주요 특성이 있는 저주파 성분이 남기 때문에 가능하다.

사람의 목소리는 고주파 성분을 제거하면 소리는 약간 다르게 들린다. 그러나, 무슨 말을 하였는지는 알 수 있다. 하지만 저주파 영역을 제거하면 알 수 없는 소리가 된다. 그런 이유로 웨이블렛 분석을 거쳐 나누어진 신호 중 저주파 부분의 근사값, 고주파 부분을 디테일이라고 한다[7-8].

IV. 구현 및 실험

4.1 실험방법

미국 콜롬비아 대학교의 LabROSA(Lab for Recognition and Organization of Speech and Audio) 연구실에서 제공하는 음성샘플을 이용한다³⁾. 잡음이 없는 상태에서 녹음된 음성에 랜덤잡음을 각 10%, 20%, 50%를 각각 혼합한 파일을 새로 생성한다. 이를 로우패스 필터한 음성과 처리전의 음성을 SNR을 비교하여, 음성의 품질을 백분율로 분석한다.

4.2 실험결과

시각적인 처리를 위해 입력된 음성신호를 시간 도메인과 주파수 도메인에서 각각 확인할 수 있도록 시뮬레이션 한다. 시간 도메인에서 음성의 크기와 음절의 단절 등의 특징을 확인할 수 있고, 주파수 도메인

에서는 입력 또는 처리된 음성의 주파수에서의 특징과 포먼트 등의 특징을 확인할 수 있다. 이를 통해 필터 등의 성능을 확인할 수 있도록 한다.

그림6은 실험을 위해 입력된 노이즈가 없는 음성샘플에 대한 시간과 주파수 도메인에서의 특성이다.

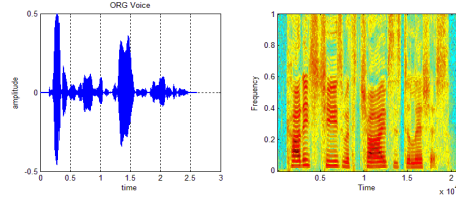


그림 6. 잡음 없는 음성신호
Fig. 6 Noise-free voice signal

그림7은 그림6의 음성신호에 각 10%의 랜덤잡음을 합성한 결과와 필터링한 결과를 각각 시간과 주파수 도메인에서의 표현하고 있다. 잡음이 포함된 음성이 필터통과 후 고역과 함께 제거되었지만, 인지하기 양호한 음성신호가 생성되고, 입력된 신호의 SNR은 12.95dB, 원 신호 대비 왜율 0.31, 상관도 99.69의 품질값을 나타냈다.

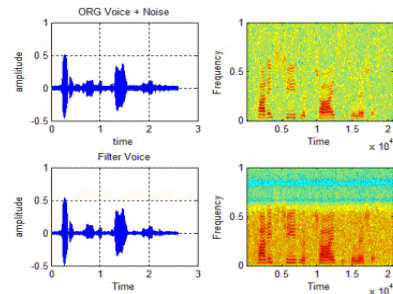


그림 7. 10% 랜덤 잡음이 포함된 음성신호
Fig. 7 The voice signal with 10% random noise

그림8은 그림6의 음성신호에 각 20%의 랜덤잡음을 합성한 결과와 필터링한 결과를 각각 시간과 주파수 도메인에서의 표현하고 있다. 그림7과 비교하여 잡음 성분이 증가하면서 음성신호와 포먼트 등이 감소된 것을 확인할 수 있다. 그러나, 필터사용으로 인지하기 양호한 음성신호가 생성되고, SNR은 10.65dB, 원 신호 대비 왜율 5.49, 상관도 94.51의 품질값을 나타냈다.

3) <http://www.ee.columbia.edu/~dpwe/sounds/>

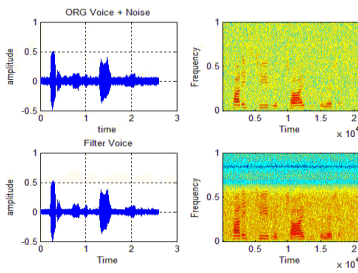


그림 8. 20% 랜덤 잡음이 포함된 음성신호
Fig. 8 The voice signal with 20% random noise

그림9는 그림6의 음성신호에 각 50%의 랜덤잡음을 합성한 결과와 필터링한 결과를 각각 시간과 주파수 도메인에서의 표현하고 있다. 그림9의 10% 랜덤잡음 합성과 비교하면 음성신호와 포먼트 등이 확연하게 감소된 것을 확인할 수 있다. SNR은 2.34dB, 원 신호 대비 왜율 61.24, 상관도 38.76의 품질값을 나타냈다.

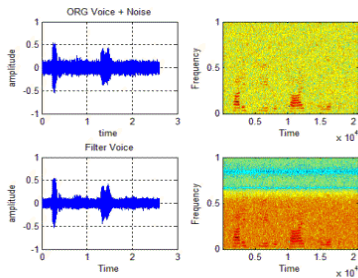


그림 9. 50% 랜덤 잡음이 포함된 음성신호
Fig. 9 The voice signal with 50% random noise

그림10은 본 실험을 위한 필터의 특성그래프를 나타낸다. 실험에 사용된 필터는 100Hz~3.4kHz의 대역을 통과시키는 로우패스 필터다. 실제로 PSTN(Public Switched Telephone Network)의 전송대역과 유사하도록 설계하였다.

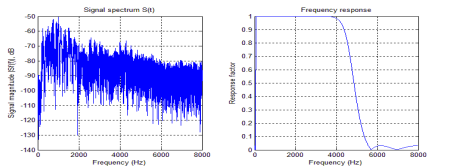


그림 10. 로우패스필터
Fig. 10 Low pass filter

최적의 음성을 선택하기 위해서 음성이 입력이 되면 VAD[9-10] 알고리즘에서 음성여부를 1차적으로 판단하고, 이를 일정한 크기의 신호로 변환하고, 잡음을 제거한 후 음성품질판별 알고리즘에서 처리 가능한 영역의 데이터로 변환한다. 이후 처리된 음성신호 중에서 최적, 최고의 음성신호를 찾아 이를 결과값으로 최적의 채널음성신호를 수신할 수 있도록 한다.

입력된 음성신호가 아주 우수하다면 품질값 90 이상의 값을 리턴하고, 음성의 품질이 아주 좋지 않다면 59 이하의 값을 리턴하여 입력된 채널 중에서 제일 높은 값을 획득한 채널의 음성을 선택할 수 있도록 한다.

본 실험과 같이 품질값 99.69, 94.51, 38.76의 3종의 음성이 입력되었다면 100에 가까운 첫 번째 음성의 품질이 제일 우수한 것이다.

음성품질판별 알고리즘에서는 최적의 음성을 판단만 하여 그 결과 값을 사용자에게 전달하여, 사용자는 이 값을 받아 수동으로 또는 가장 높은 값을 자동으로 최적 채널을 선택할 수 있는 기준값을 갖도록 한다. 이는 후에 최적 음성채널 선택을 할 때 사용자의 의견을 반영하여 출력하기 위함이다.

V. 결 론

MOS알고리즘은 유선전화, 이동통신 등에서 음성의 품질을 판단하는 가장 오래되고, 많이 사용되는 알고리즘으로 음성처리, 음성인식 등에서 전처리에 많이 사용되는 안정적이고 단순한 알고리즘이다. MOS는 5단계로 품질을 구분하나, 이를 더 세분화하여 100분율에 해당하는 음성품질 데이터를 사용자에게 전달함으로써, 기존에 단순히 음량의 크기를 품질로 판단하는 경우에 비해 향상된 품질의 음성채널 선택할 수 있다. 이를 항공기 관제나 훈련감청 시스템 운영에 활용하면, 잡음환경에서도 현재 보다 좋은 품질의 음성채널을 선택할 수 있어 안전사고 예방과 훈련 능력향상 등을 기대할 수 있다.

References

[1] S. Park and S. Ryu, "Quality Measure Over Mobile Communication Using Bark Coherence Function," *J. of the Korea Institute of Communication Sciences*, vol. 26, no. 4, Apr. 2001, pp. 437-446.

[2] B. Kim, "Service Quality Criteria for Voice Services over a HSDPA System," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 7 no. 2, Apr. 2012, pp. 249-255.

[3] B. Kim and G. Cho, "An Intergated E-model Implementation for Speech Quality Measurement in VoIP and VoLTE," *J. of the Institute of Electronics and Information Engineers*, vol. 50, no. 7, July 2013, pp. 10-18.

[4] S. Lee and K. Sung, "Objective Speech Quality Assessment Technique," *J. of the Korean Institute of Communication Sciences*, vol. 22, no. 10, 2005, pp. 24-34.

[5] ITU-T Recommendation G.107, "The E-Model : a computational Model for use in Transmission Planning," *International Telecommunications Union Telecommunication*, Ottawa, Ont., Canda, Oct. 2008.

[6] Y. Kim and H. Lee, "A Study on Improved Method of Voice Recognition Rate," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 8 no. 1, Jan. 2013, pp. 77-83.

[7] S. Jung and J. Hong, "A Study on the Speech Blind Source Separation Using Wavelet Transform and Independent," *J. of the Institute of Electornics Engineers of Korea*, vol. 40, no. 2, June 2003, pp. 15-22

[8] C. Lee and D. Kim, "Adaptive Noise Reduction of Speech Using Wavelet Transform," *J. of the Korea Institute of Electronic Communication Sciences*, vol. 4 no. 3, Sept. 2009, pp. 190-196.

[9] Y. Park and S. Lee, "Voice Activity Detection Using Global Speech Absence Probability Based on Teager Energy in Noisy Environments," *J. of the Institute of Electronics Engineers of Korea*, vol. 49, no. 1, Jan. 2012, pp. 97-103.

[10] Y. Chung, "An Optimal SNR Model Selection Method in the Multi-Model Based Speech Recognizer," *J. of Korean Institute Of*

Information Technology, vol. 9, no. 7, July. 2011, pp. 65-70.

저자 소개

류창국(Chang-Guk Lew)



1991년 청주대학교 전자공학과 졸업(공학사)

2001년 한국항공대학교 항공산업 대학원 정보통신공학과 졸업(공학석사)

2014년 ~ 현재 전남대학교 대학원 전자컴퓨터공학과 박사과정

※ 관심분야 : 신호처리, ITS, u-City, 국방통신시스템

이배호(Bae-Ho Lee)



1978년 한양대학교 전자학과 졸업(공학사)

1980년 한국과학기술원 대학원 전기 및 전자공학과 졸업(공학석사)

1993년 University of Missouri, Columbia 졸업(공학박사)

1980년 ~ 1983년 국방과학연구소

1993년 ~ 현재 전남대학교 전자컴퓨터공학부 교수

※ 관심분야 : 영상처리, 컴퓨터비전, 인공지능, 멀티미디어 등

