로그인 과정에서의 화자인증 메커니즘을 이용한 사용자인증 방안 연구

(A study on user authentication method using speaker authentication mechanism in login process)

김남호*, 최지영**
(Nam-Ho Kim, Ji-Young Choi)

요 약

인터넷과 스마트폰 활용이 대중화되면서 사용자들은 다양한 방법과 미디어를 통해 언제 어디서나 정보시스템에 접근하여 필요한 서비스를 이용하는 다채널 환경에서 살고 있다. 이러한 서비스를 받는 과정에서 사용자는 본인임을 인증하는 사용자 인증 단계를 필수적으로 수행해야 하는데 대표적인 방식이 아이디 패스워드 인증 방식이다. 아이디 패스워드 기반의 사용자 인증 방식은 본인의 인증정보를 기억 후 키보드를 통한 입력만으로 인증이 가능하므로 타 인증 방식들과 비교했을 시 가장 편리하다는 평가를 받는다. 반면 현대 웹 서비스에선 요구하는 패스워드의 문자열 조합방식이 다르고 복잡성이 높은 엔트로피 값의 패스워드 설정만을 허용하고 있다. 이러한 복잡한 문자열로 구성된 패스워드는 사용자가 이용하고자 하는 서비스가 많을수록 개인정보 유출방지를 위해 주기적으로 패스워드 변경을 권고하기 때문에 이를 기억해야 할 사용자 인증정보의 개수도 비례하여 증가한다. 이러한 높은 엔트로피 값을 가지는 사용자의 인증정보를 시각 장애인이나 손사용이 불편한 사람 혹은 고령층이 일일이 기억하고 키보드 입력을 통해 사용하기엔 어려움이 따른다. 따라서 본 논문에서는 위와 같은 취약계층 및 일반 사용자에게 로그인 과정에서의 간편한 사용자 인증 방식 제공을 위해 구글 어시스턴트와 MFCC 및 DTW 알고리즘 그리고 화자 인증을 사용한 사용자 인증 방식을 제안한다.

■ 중심어 : 화자 인증 ; MFCC ; 사용자 인증 ; DTW ; 구글

Abstract

With the popularization of the Internet and smartphone uses, people in the modern era are living in a multi-channel environment in which they access the information system freely through various methods and media. In the process of utilizing such services, users must authenticate themselves, the typical of which is ID & password authentication. It is considered the most convenient method as it can be authenticated only through the keyboard after remembering its own credentials. On the other hand, modern web services only allow passwords to be set with high complexity by different combinations. Passwords consisting of these complex strings also increase proportionally, since the more services users want to use, the more user authentication information they need to remember is recommended periodically to prevent personal information leakage. It is difficult for the blind, the disabled, or the elderly to remember the authentication information of users with such high entropy values and to use it through keyboard input. Therefore, this paper proposes a user authentication method using Google Assistant, MFCC and DTW algorithms and speaker authentication to provide the handicapped users with an easy user authentication method in the login process.

■ keywords : Speaker Authentication ; MFCC ; User Authentication ; DTW ; Google

Ⅰ. 서 론

오늘날 IT 기술력의 발달로 인터넷과 스마트폰 기기의 보급이 활성화 되고 사용자들은 다양한 방법과 매체를 통해 언제 어디서나 정보시스템에 접근하여 원하는 서비스를 제공받을 수 있는 환경에서 살고 있다. 이러한 과정에서 사용자들은 서비스를 사용하기 위해 본인임을 인증하는 사용자 인증과정을 필수적으로 거쳐야 한다. 일상생활에서 가장 많이 사용되는 인증방식은 지식 기반 인증(something you know) 방식 중 하나인

* 정회원, 호남대학교 대학원 소프트웨어공학과 졸업
** 정회원, 호남대학교 소프트웨어학과

접수일자 : 2019년 06월 05일
수정일자 : 1차 2019년 07월 03일
게재확정일 : 2019년 07월 03일
교신저자 : 김남호, e-mail : nhkim@honam.ac.kr

아이디 패스워드 사용자 인증 방식이 있다. 하지만 아이디 패스워드 기반의 사용자 인증 방식은 사전에 본인이 이용하고자 하는 서비스에 접근하여 아이디와 패스워드를 회원등록과정에서 등록 후 이를 기억하고 있어야 한다. 하지만 이러한 인증 방식은 사용자의 개인정보 유출을 방지하기 위해 패스워드를 주기적으로 변경할 것을 권고하고 있으며 문자열 조합방식에 따라서 낮은 엔트로피 값의 패스워드는 사용하지 못하도록 특수문자 및 영문 대소문자 숫자를 조합하여 복잡성을 높인 특정 이상의 엔트로피 값 이상일 때 패스워드 설정을 허용하고 있다. 아울러 웹사이트마다 패스워드를 요구하는 문자열 조합방식도 조금씩 차이가 있기에 제공 받고자 하는 서비스가 많을수록 사용자가 기억하고 있어야 할 아이디와 패스워드는 서비스의 개수에 따라 비례하여 증가한다. 이는 일반적인 사용자에게도 번거로운 일이지만 키보드 입력에 제약을 받는 시각 장애인이나 손 사용이 불편한 사람이나 고령층의 시각에서 본다면 아이디와 복잡한 패스워드를 항시 기억하고 입력해야 하는 과정은 상당히 불편하다. 아울러 시각 장애인을 위한 점자 키보드 및 보조장비 같은 경우 매우 고가의 기기이므로 이는 장애인의 소득과 생활여건을 고려해보았을 때, 개인이 구비하기가 매우 어렵고 PC방 카페 등 공공장소 데스크탑 같은 경우 이처럼 시력이 감퇴한 시각 장애인이 급한 상황에 PC를 이용하고자 하여도 이에 따른 보조 장비를 항시 소지하고 있지 않기 때문에 사용하기가 어렵다[1]. 따라서 본 논문에서는 위와 같은 웹서비스 사용자 인증과정에서 일반적인 아이디 패스워드 기반의 사용자 인증 방식보다 시각 장애인과 같은 키보드 입력이 취약한 사람들이 기존의 인증 방식보다 음성을 이용한 STT(Speech to Text), TTS(Text to Speech), 화자인증만으로 인증 가능한 항시 소지할 수 있는 스마트폰 기기를 활용하여 웹사이트에 기존에 회원등록이 되어있는 사용자들을 대상으로 MFCC(Mel-Frequency Cepstral Coefficient)와 DTW (Dynamic Time Warping) 알고리즘을 사용하여 화자인증을 구현하였으며, 키 입력을 사용하지 않고도 간편하게 사용자 인증을 수행할 수 있는 시스템을 제안하고자 한다.

Ⅱ. 관련 연구

1. 구글 어시스턴트

구글 어시스턴트(Google Assistant)는 2016년 5월 18일 미국 캘리포니아 마운틴뷰에서 열린 구글 개발자회의(Google I/O)에서 공개한 구글이 개발한 아이폰에서 제공하는 시리와 비슷한 기능을 수행하지만 이와 비교해 좀 더 다양한 패턴의 질문을 할 수 있고 상대적으로 인식률이 높은 인공지능(AI) 비서 시스템이다. 구글 어시스턴트는 한국어를 지원하며 이의 동작 방식은 사용자의 음성을 인식 후 사용자의 질문을 파악하여 앱 어플리케이션 실행, 음악 재생, 웹사이트 검색, 날씨 조회, 메시지 전송 등 다양한 기능을 수행할 수 있으므로 시력이 감퇴한 시각 장애인이나 고령층이 편리하게 사용하기에도 적합하다. 아울러 구글 어시스턴트를 실행시키는 방식도 스마트폰 기기의 홈버튼을 짧은 몇 초 동안 누르고 있으면 실행시킬 수 있어 간편하다[2].

그림 1. 구글사의 인공지능 비서 시스템

2. MFCC

MFCC(Mel-Frequency Cepstral Coefficient)는 MFC의 계수들의 집합이다. 이는 입력된 음성 신호를 인간의 청각기관으로 모델링 하여 변환하는 음성의 특징 추출의 한 형태를 말하며 음성 신호처리에서 LPC(Linear Prediction Coefficients)나 LPS(Linear Prediction Spectrum)과 같이 대표적으로 음성 신호에서 특징 추출 방법의 하나로 LPC나 LPS와 비교해서 채널 왜곡이나 주변 잡음에 강하므로 인식 성능이 좋은 것으로 알려져 있다[3]. 또한 MFCC는 사람 음성의 특징 추출 뿐 아니라 다방면의 신호에 관련된 신호처리에서 가장 많이 사용되고 있는 방식 중 하나이다. 대표적으로 MFCC는 최근에 GMM(Gaussian mixture model)을 이용한 화자 인식에 많이 사용되고 있다. 이에 해당하는 MFCC의 특징 벡터 추출 순서는 다음 그림 2와 같은 순서로 추출된다.

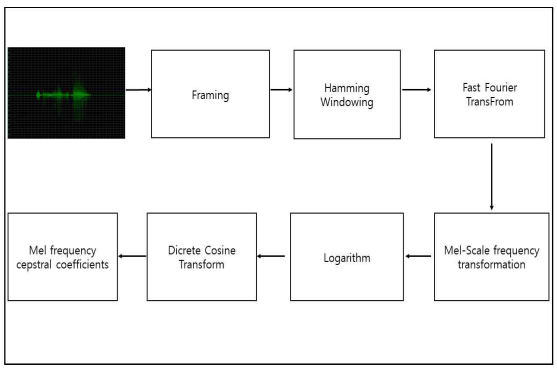


그림 2. MFCC 특징 벡터 추출 순서도

가장 먼저 음성 신호는 A/D 변환을 거친 후 디지털신호로 변환되고 이후 Pre-Emphasis 필터 과정을 거치게 된다. 이는 인간의 고주파수 대역 강조 필터로서 유성음 부분은 생리적인 특

성 때문에 20db 정도 감쇄하는 것을 보상하여 음성으로부터 성도만의 특성을 얻는데 사용된다. 이후 Framing 과정은 일정한 간격으로 음성 신호를 자르는 역할을 하며 20ms에서 40ms 가량의 단위 시간으로 프레임을 자른 후 일정한 간격으로 자른 프레임 크기의 절반에 해당하는 크기만큼 중첩 시키며 이동한다. 이후 중첩 시키며 일정한 간격으로 나누어진 프레임에 각각 해밍 윈도우(Hamming Window)를 적용 시킨 후 이에 해당하는 음성 신호에 대한 파워스펙트럼을 청각기의 주파수 반응도를 모사한 mel-Scale 주파수 도메인에서 DCT(Discrete Cosine Transform)를 적용하여 저주파 계수 중 원하는 차수만큼 MFCC의 계수를 추출할 수 있게 된다. 보통 사용되는 계수의 차수는 12차에서 19차까지 주로 사용되며, 위의 그림 2처럼 MFCC의 추출 과정은 총 7단계를 거치게 된다. MFCC는 한 개의 프레임 안에 해당하는 음성 데이터의 여러 차수의 계수를 추출함으로써 이를 음성 신호의 특징 벡터로 활용하게 된다. 본 논문에서는 이에 사용되는 MFCC 계수를 프레임마다 12차의 계수를 추출하였으며 여기에 Logarithm 과정을 거친 필터 뱅크의 개수만큼 추출한 에너지의 합을 포함하여 사용하였다[3, 4, 7].

3. DTW

DTW(Dynamic Time Warping) 알고리즘은 길이나 속도가 다른 기준 패턴과 참조패턴 두 개의 시계열 데이터 사이에서 최적의 정합 경로를 찾아 유사도를 측정하여 오차 거리를 최소화 하는 동적 프로그래밍(Dynamic Programming) 기반의 비교적 간단한 시스템으로 좋은 성능을 얻을 수 있는 알고리즘이다[10]. 음성 신호는 사람마다 발성 및 습관에 따라 발성 속도가 다르고 같은 단어를 발성하여도 그림 3의 A와 B와 같이 단어적 시간의 길이가 변화하기 때문에 기존의 거리 비교 알고리즘으로 계산하면 같은 지점의 거리를 계산하기 때문에 시간 축이 고르지 않아 오인식이 발생하므로 이와 같은 시계열 데이터 비교에선 DTW 알고리즘이 사용된다.

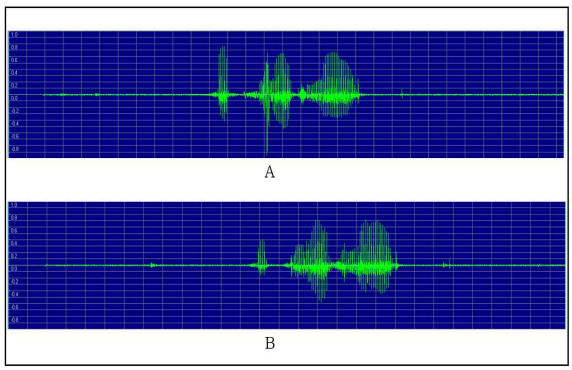


그림 3. 같은 단어 발성 시 음성 신호의 길이 차이

DTW를 이용한 시계열 패턴 비교 방식은 다음과 같다. 입력 패턴의 길이가 L인 기준 패턴을 A=a(1),a(2),a(3),...,a(L)이라 하고 입력 패턴의 길이가 M인 참조패턴을 B=b(1),b(2),b(3),b(4),...,b(M) 이라고 할 때 두 개의 상이한 신호 패턴의 유사도 C는 수식 1과 같이 표현되며 누적 거리 C를 최소화하여 최적 경로를 찾는 방법이다[4].

$$C = \sum_{n=1}^{A} d(A(n), B(w(n))) \quad (1)$$

위와 같은 최적 경로를 탐색할 때는 다음과 같은 4개의 제약조건을 더해 탐색 시간을 줄이게 된다[3].

1. 시작과 끝점 제약(endPoint constraints)
2. 단조 증가조건(monotonically increasing)
3. 전역 경로 제약(global path constraints)
4. 국부 경로 제약(local path constraints)

III. 화자인증 등록과정 및 인증과정

1. 화자 인증을 위한 단어 선정

본 화자인증시스템은 사용자 인증을 하는 과정에서 사용자의 음성을 입력하여 등록하는 단계와 입력한 음성을 통한 사용자 인증 단계로 구성된다. 음성의 특징 추출은 화자가 발성한 시간의 흐름에 따라 변하는 음성 신호에서 MFCC(Mel-Frequency Cepstral Coefficient) 특징 벡터를 이용하여 각각 분리한 프레임마다 총 12차의 MFCC 계수를 추출하였으며 DTW(Dynamic Time Warping) 알고리즘을 통해 비교 연산을 진행하도록 설계하였다. 또한 화자가 발성을 하는 과정에서 단어의 선택이 중요하다고 알려져 있다. 따라서 본 논문에서는 화자 인증과정을 수행하는 한국어 자음 발음의 단어 선택을 자의적인 기준에 의해 분류하지 않았으며 음가를 결정짓는 요인들을 기준으로 분류하여 화자인증을 설계 및 구현하였다. 본 논문에서 아래의 그림 4에 해당하는 모든 단어의 목록을 사용하진 않았으며 위의 단어 중 화자인증에 주로 사용된 단어는 과자, 할머니, 기자, 거북이 등 편의성을 위해 더욱이 짧은 단어를 위주로 사용하였다[5].

No.	단어	No.	단어	No.	단어	No.	단어
1	하품하지마	14	호리병	27	형편	40	바글대다
2	허비하지마	15	호미	28	혁명	41	함구무언
3	할아버지	16	호박	29	허벅다리	42	바구니
4	할머니	17	혼비백산	30	헌병	43	화병
5	학부모	18	화분	31	헌법	44	화분
6	하마	19	화장품	32	허파	45	호랑나비
7	홍미	20	자동차	33	허비	46	홀가분하다
8	홍분	21	회오리바람	34	학벌	47	바가지
9	흑백영화	22	자르다	35	학도병	48	혹부리
10	후백제	23	가제걸음	36	학비	49	형부
11	후보	24	형무소	37	기자	50	화살표
12	자동문	25	형벌	38	한문		
13	효모	26	과자	39	한복		

그림 4. 화자인증에 사용된 단어 목록[5]

2. 화자인증 등록과정

아래의 그림 5에선 본 논문에서 설계한 화자 인증시스템의 전체적인 구조이다. 화자인증의 녹음 설정은 16000hz Sampling rate와 Mono 단일 채널로 구현하였으며 프레임 크기는 30ms 가량의 시간인 512만큼 나누었다. 512만큼 프레임을 이동시키면서 중첩 구간은 프레임의 절반 크기인 256으로 설정하였다. 이때 프레임의 크기가 너무 작으면 주파수의 해상도가 낮아지고 프레임의 크기가 클수록 주파수의 해상도가 높아지는데 이는 정밀도에 있어서 영향을 줄 수 있으므로 적합한 프레임의 크기 설정이 중요하다.

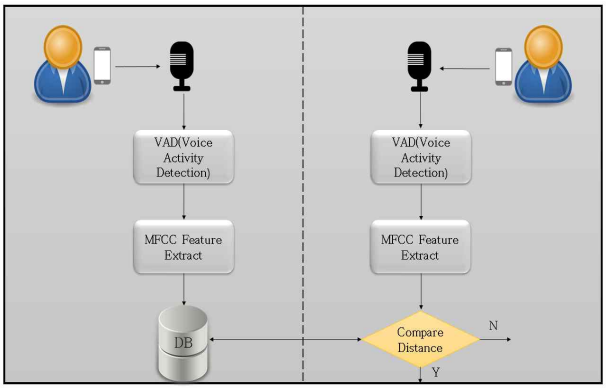


그림 5. 화자인증 절차도

이에 해당하는 녹음의 환경은 30db에서 40db 정도의 소음 크기인 다소 조용한 공간인 일반적인 가정집 소음환경에서 진행하였다. 사용자가 목소리를 발성 후 무손실 무압축 방식인 wav 파일 형식으로 저장하였다. 이처럼 저장된 wav 음원 파일에서 사람의 음성이라고 판단되는 구간만 획득하는 VAD(Voice Activity Detection)방식은 저장된 음성 신호들을 프레임 단위로 에너지를 구한 후 특정 에너지 이상으로 신호가 검출될 때까지 무음 구간의 프레임을 제거하였으며 위와 같이 무음 구간이 제거된 음성 신호에서 MFCC의 특징 벡터를 추출하였다. 이러한 음성 구간 검출의 중요성은 화자 인식에 있어서 정확도와 관계가 있다. 음성구간 검출을 함으로서 음성이 존재하는 구간만 비교하여 인식률을 상승시킬 수 있으며 음성 특징 추출 과정에서 밀리세컨드 단위로 나누어진 프레임마다 MFCC 계수를 추출하기에 많은 연산을 필요로 하게 되는데 음성이 존재하지 않는 구간은 버려지므로 음성 특징 추출 연산 속도도 향상 시킬 수 있다[10].

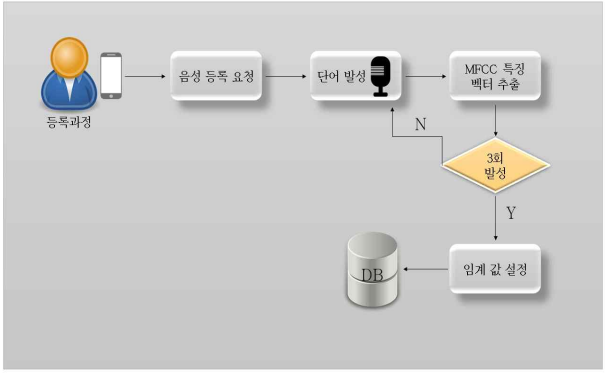


그림 6. 화자인증 등록과정

그림 6은 그림5의 화자 인증 전체 순서도의 등록과정에 해당하는 그림이다. 본 등록과정에서는 총 3번의 발성이 이루어지며 각각의 음원 파일에서 사람의 목소리의 구간을 획득하고 MFCC의 특징 벡터를 추출한다. 3개의 음원 파일에서 추출한 특징 벡터를 각각 그림 7과 같이 DTW 알고리즘으로 비교하여 3개의 평균 Distance 값을 구하였다. 이에 더해 평균 Distance의 값은 구해진 평균의 거리 값을 그대로 사용하지 않고 10~20% 정도의 민감도 값을 별도로 부여하여 이를 화자 인식의 거리 임계값으로 설정하였다. 이때 DTW 알고리즘은 시계열 데이터 간의 비교를 통한 결과는 참조패턴과 기준 패턴이 유사할수록 수치가 0에 가깝게 나오는 알고리즘이므로 등록과정의 3번의 발성 과정에서 Distance 값의 평균이 너무 낮게 나오게 된다면 인증과정 중 등록과정에서 설정한 임계 값 이하여야만 사용자 인증이 성공하기 때문에 인식률 상승을 위해 별도의 민감도 값을 부여하였다.

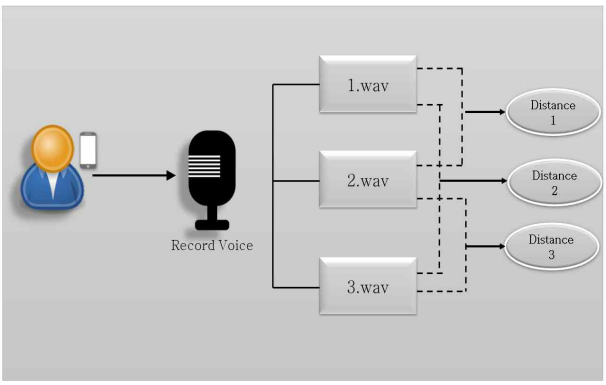


그림 7. 화자인증 임계값 설정과정

$$Threshold = \frac{d1 + d2 + d3}{n} \quad (1)$$

$$Threshold \mathrel{+}= Threshold * 0.2 \quad (2)$$

3. 화자인증 인증과정

본 단계는 화자인증 등록 후 사용자가 인증과정을 이용할 시 이루어지는 화자인증 단계로서 인증을 위한 새로운 음성 녹음 후에 화자 등록과정에서 그림 7의 1.wav, 2.wav, 3.wav에서 추출한 MFCC 특징 벡터값들과 각각 일대일 매칭 비교하여 DTW 알고리즘을 통해 비교 연산을 진행한다.

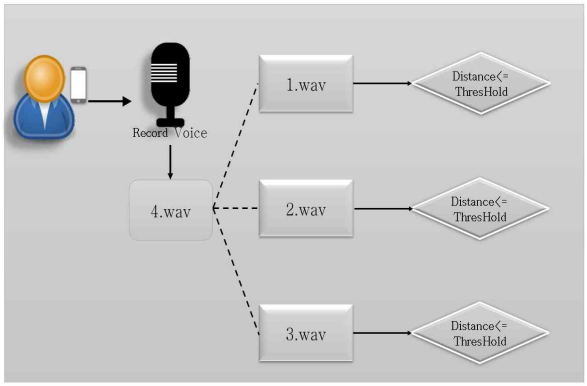


그림 8. 화자인증 인증과정

이후 3개의 Distance 값을 얻으며 이 중 한 개 이상의 값이 임계값으로 설정해놓은 Distance 수치 미만으로 나오면 해당 사용자로 인식하여 승인과정이 이루어진다. 이러한 인증 승인과정 중 두 개 이상의 값이 임계 값 미만이 되었을 때를 인증 승인 기준으로 설정하였을 때 한 개 이상의 값이 임계값 이하를 인증 승인 기준으로 설정했을 때보다 본인 거부율의 비율 즉 FRR(False Rejection Rate) 의 비율이 높아진 결과를 볼 수 있었으므로 본 논문에서는 화자인증 승인 기준을 한 개 이상의 값이 임계값 이하인 상황을 인증 성공 기준으로 설정하였다.

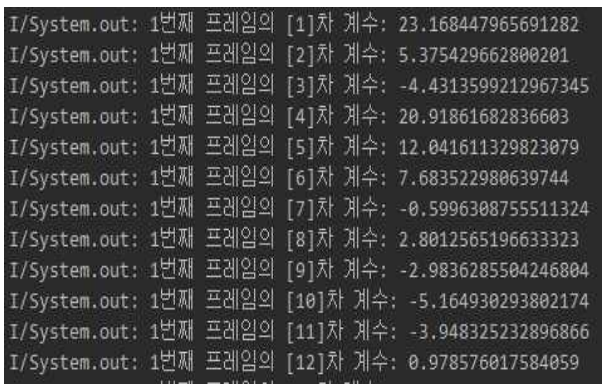


그림 9. 12차 MFCC 계수 추출 화면

그림 9는 입력된 음성의 MFCC 특징 벡터를 추출한 값이며 위와 같은 방식으로 화자 1이 '과자'라는 단어를 3회 발성한 후 이에 해당하는 평균 임계값이 30.00이 나왔을 때 위의 수치에 20% 정도의 만감도 값을 부여하여 36.00이 설정되게 하여 진행하였다. 임계 값 결정 후 화자 2(남성)과 화자 3(여성)이 같은 단어인 과자를 15번씩 발성하였을 때 평균 결과는 61.09, 71.26이 나왔으며 본인과 타인이 동일한 단어를 발성했을 시 30.00 이상의 표준편차가 나타남을 볼 수 있었고 이외에도 다른 화자가 같은 단어인 '과자'를 발성했을 때도 비슷한 수치 이상의 차이를 보였으므로 음성을 등록한 본인이 아닌 타인은 인증을 수행할 수 없다. 아울러 타인은 사전에 화자가 등록한 단어를 모르므로 인증이 진행되지 않아 해당 발성하는 단어 자체가 1차 보안이 될 수 있고 전송한 음성 특징 정보가 임계값을 통해 인증 여부가 결정되므로 이를 2차 보안으로 볼 수 있다.

Ⅳ. 화자 인증 설계를 기반으로 하는 사용자 인증 시뮬레이션

본 논문에서 제안하는 인증시스템의 구성요소의 첫 번째는 사용자가 서비스를 사용하기 위해 사용자 등록 및 사용자 인증요청에 해당하는 웹 어플리케이션(Web application), 두 번째는 사용자의 음성을 전송받아 임계 값 비교를 통해 승인 혹은 거부 인가 여부를 결정짓는 웹 어플리케이션 서버(Web application Server), 세 번째는 웹서버에서 등록 및 인증이 허가된 사용자의 인증정보를 저장하기 위한 데이터베이스(Data Base), 네 번째 본인의 음성을 녹음 후 음성의 특징 벡터를 추출하여 전송하는 스마트폰 기기 즉 앱 어플리케이션 (App application)으로 크게 네 가지로 분류된다.

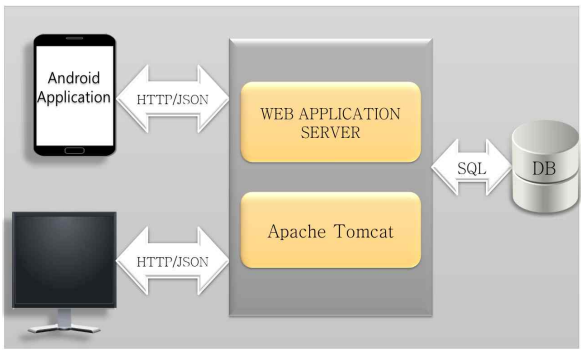


그림 10. 시스템 구성도

또한 본 단계에서는 제안한 화자 인증시스템을 기반으로 기존의 웹사이트에도 적용할 수 있게 사전에 웹사이트에 회원가입이 되어있는 사용자들을 대상으로 하였다. 시각 장애인이나 고령층과 같이 시력이 일반 사람들보다 약시인 사용자를 위한 웹사이트에서의 사람의 음성정보를 통한 사용자 인증 방안을 검증하였다.

가장 먼저 사용자가 웹 서비스에 접근하여 화자인증 시스템을 사용하기에 앞서 서버를 통한 화자인증을 위해 단어 발성을 통해 본인의 음성을 등록하여야 한다. 이와 같은 등록을 위해 스마트폰 기기의 마이크를 통해서 음성 녹음을 진행한다. 웹서버에 등록 요청을 진행하면 웹서버는 요청이 들어온 사용자의 PC 화면에 현재 사용 중인 PC와 연결되어있는 SessionID를 통해 이를 중복되지 않는 다음 그림 11의 'ABCD'와 같은 단순한 패턴의 숫자나 문자, 즉 짧은 숫자나 짧은 문자열로 재구성하여 현재 사용 중인 PC에 임시 ID를 부여한다. 이는 시각 장애인들과 같이 약시인 사람들에게 편의성 제공을 위함이며 스마트폰 기기 즉 App에서 인증정보를 전송했을 시 사용자가 현재 사용 중인 PC를 식별하기 위한 정보로 활용된다.

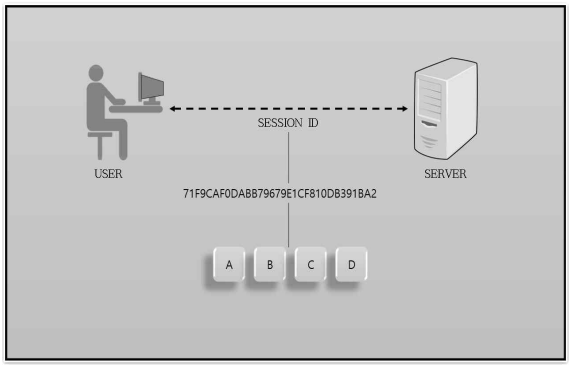


그림 11. PC를 식별하기 위한 정보

아래의 그림 12와 같이 사용자의 PC 화면에 짧은 문자열로 변환된 SessionID를 텍스트를 음성으로 변환시키는 TTS(Text to Speech)를 통해 사용자의 스피커로 출력한다. 사용자는 'ABCD' 같은 짧은 변환된 문자열을 들은 후 사용자 등록을 위해 본인의 스마트폰 기기를 통해 구글 어시스턴트 인공지능 App을 실행시킨다. 예를 들어 사용자가 실행시킬 앱 어플리케이션의 이름이 '보이스 인증' 이라면 '보이스 인증 열어줘' 혹은 '보이스 인증 실행'이라는 음성 명령어를 통해 해당 앱을 실행시킨다.

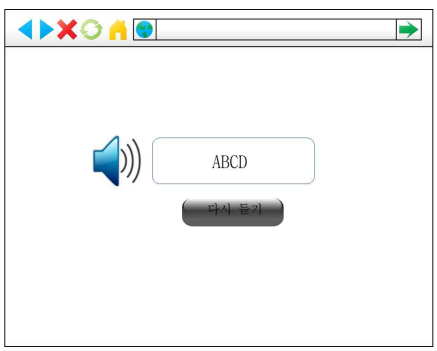


그림 12. 변환된 SessionID TTS(Text to Speech)

위와 같은 과정에서도 사용자는 음성으로 App의 이름만 알려주면 App은 자동으로 실행되므로 별도의 터치나 입력 값은 필요하지 않다. 이에 해당하는 구글 어시스턴트를 통한 App의 실행 예시는 다음 그림 13과 같다.

그림 13. 구글 어시스턴트를 통한 App실행 예시

App 실행 후 사용자는 전체 영역에 해당 되는 Activity를 클릭하여 사용자가 선행 과정에서 들었던 변환된 문자열을 스마트폰 기기 마이크에 발성한다. 이러한 과정에서 음성을 텍스트로 변환하는 STT(Speech to Text)를 통해 음성을 다시 텍스트로 바꾸어 이를 서버에서 PC를 식별하기 위한 정보로 사용한다. 이후 스마트폰 기기 내부적으로 국제 스마트폰 기기 고유 식별 번호인 IMEI값을 불러온다. IMEI 번호는 휴대 전화마다 부여되는 기기의 고유번호로서 세계 이동통신 사업자연합(GSMA)의 가이드라인에 따라 삼성전자나 모토로라와 같은 제조업체에서 부여한 값이며 고유번호이기 때문에 중복되지 않는 고유한 숫자이다[9]. 이후 Activity의 Intent를 통한 화면 전환이나 Activity Dialog를 통해 사용자의 음성을 등록하기 위한 화면으로 이동한다. 예를 들어 화자인증 등록과정에서 사용자가 스마트폰 마이크에 '과자'라는 단어로 등록할 시 이에 해당하는 단어를 총 3번의 발성하는 과정이 이루어지고 각각의 MFCC의 특징 벡터를 추출 후 특징 벡터와 IMEI 그리고 세션 아이디를 암호화하여 서버로 전송한다.

그림 14. 서버 전송을 위해 암호화한 MFCC 특징 벡터

MFCC 특징 벡터를 전송받은 서버는 'ABCD' 에 해당되는 PC를 식별 후 3회 녹음된 음원 파일의 MFCC 특징 벡터의 임계값을 구하고 IMEI 번호를 DB 인증 서버를 통해 해당 사용자의 정보를 저장 후 등록이 완료된다. 이와 같이 사용자의 음성을 등록 후 화자인증을 통한 사용자 인증과정은 등록과정과 동일한 방식으로 진행된다. 사용자가 사용하고자 하는 PC를 통해 웹사이트에 접속하여 인증요청하면 SessionID를 변환하여 사용자 PC에 출력 후 이를 스피커를 통해 사용자에게 알려주고 사용자는 구글 어시스턴트를 통해 인증을 사용하려는 App을 음성 명령어를 발성하여 실행한다. 이후 본인이 사전에 등록과정에서 등록한 단어를 발성 후 서버로 사용자를 식별할 수 있는 IMEI 번호와 변환된 세션 값 그리고 MFCC 특징 벡터를 전송하여 인증 여부를 수신한다.

III. 결 론

본 논문에서는 위에서 구현한 화자인증 시스템 및 구글 어시스턴트 App과 음성 합성 기술을 종합하여 웹 기반 사용자 인증 시뮬레이션에 적용해 봄으로써 약시자나 손사용이 불편한 사람과 같이 취약계층뿐 아니라 일반 사용자에게도 사람의 음성만을 통한 간편한 사용자 인증이 될 수 있는 인증시스템을 설계 및 제안하였다. 화자인증의 설계 및 구현은 스마트폰 기기나 PC 환경 어디에도 적용할 수 있는 JAVA언어를 사용하여 MFCC와 DTW를 구현하였고, 이러한 설계 및 구현 과정 중 등록과정에서 이를 사용하고자 하는 화자는 본인의 음성을 세 번 등록하게 되는 이와 같은 과정에서 별도의 민감도 값을 부여하여 임계값을 설정하는 방식으로 진행하였다. 이러한 임계값 설정 시 DTW 알고리즘을 통한 거리 값의 평균은 필터 뱅크의 개수나 프레임의 크기 등 설정의 요인에 따라 평균 임계값이 다르게 나타나고 동일 설정으로 진행하였을 때 동일 화자가 여러 번 발성하는 경우 비슷한 Distance의 수치가 나타나는 것을 볼 수 있었으며, 본인이 등록한 음성에 타인이 동일 단어를 발성하더라도 일정 이상의 표준편차 Distance 수치 차이가 나는 것을 볼 수 있었다. 하지만 사용자 인증과정에서 등록한 사용자의 음정 변화가 없는 상황에는 일정한 인증 성공율을 보였지만 감기에 걸렸을 경우 혹은 음성의 크기 등 음정의 변화와 장소의 이동 등 즉, 환경의 변인에 영향을 받아 상대적으로 본인 거부율(FRR)의 비율이 높아지는 걸 볼 수 있었다. 따라서 향후 연구에서는 이러한 환경적인 변인 요인이 존재하는 상황 속에서도 화자를 식별하여 인증할 수 있는 알고리즘의 연구와 임계값 설정 부분에 있어서 본 논문에서 제안한 임계값 설정 방식보다 조금 더 세밀하고 정확한 임계값 설정에 대한 연구가 보완된다면 본인 거부율의 비율을 낮추면서 웹 기반 서비스뿐 아니라 사용자에게 본인 인증 과정이 필요한 어느 매체에서나 모두 적용이 가능할 것으로 기대된다.

REFERENCES

[1] J.G. Chae, J.W. Jang, D.W. Kim, S.J. Jung, and I. H. Lee, "Voice Assistant for Visually Impaired People," *The Journal of Korean Institute of Information Technology*, vol. 17, no. 4, pp. 131-136, Apr. 2019.

[2] 구글 어시스턴트 sdk. https://developers.google.com/assistant/sdk/ (accessed May, 22, 2019).

[3] J.H. Jo, H.Y Yoo, S.Y Cha, and I.C. Park, "Optimization of Floating-Point Bit-width for MFCC Feature Extraction," *대한전자공학회하계학술대회*, 제36권, 제1호, 1194-1197쪽, 2013년 7월

[4] S.D. Jeong "Speaker Identification Using Dynamic Time Warping Algorithm," *한국산학기술학회 논문지*, 제12권, 제5호, 2402-2409쪽, 2011년 5월

[5] 강현종, "화자인증 기술을 이용한 웹 기반 출석 체크 시스템 구현", *단국대학교 석사학위 논문*, 2007. 8

[6] Android SpeechRecognizer. https://developer.android.com/reference/android/speech/SpeechRecognizer (accessed May, 22, 2019).

[7] 장한, 김학태, 정길도. "청각 주파수 응답에 기반한 자동 모음 개시 지점 탐지," *한국산학기술학회 논문지*, 제13권, 제1호, 333-342쪽, 2012년 1월

[8] 이진우, 김선주, 조인준, "USIM 정보를 이용한 사용자 인증 방안 설계 및 구현," *한국콘텐츠학회논문지*, 제17권, 제7호, 571-578쪽, 2017년 7월

[9] 국제모바일기기 식별코드. https://terms.naver.com/entry.nhn?docId=973170&cid=43667&categoryId=43667 (accessed May, 22, 2019).

[10] J.Y. Kim, S.J. Baek, and Y.K. Nam. "The Implementation of Speaker Verification Tool using GMM(Gaussian Mixture Models)," *전기통신기술연구센터*, 제6권, 제1호, 46-52쪽, 2003년 12월

[11] 최지영, 김남호, "QR코드 로그인 과정에서의 본인 확인을 위한 화자인증 메커니즘," *한국스마트미디어학회 2019춘계학술대회논문지*, 2019년 4월

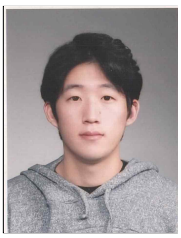
[12] L. Muda, M. Begam, and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques," a*rXiv preprint arXiv*:1003.4083, Mar 2010.

저 자 소 개

김남호(종신회원)

1997년 포항공과대학교 정보통신학과 석사 졸업
2013년 전남대학교 전산통계 박사 졸업
1991년~1997년 포스코ICT(주) 연구원
1998년~현재 호남대학교 소프트웨어학과 부교수

<주관심분야 : 사물인터넷, 인공지능, 응용 SW>



최지영(정회원)

2017년~2019년 호남대학교 대학원 소프트웨어공학과 석사 졸업

<주관심분야 : 미디어처리, 정보보안 시스템 보안>