이동단말에서 다중발화율 이용한 Home network 환경에서의
QoS 보장 연구*
황지수○ 이창섭 박준석 김유섭 박찬영
한림대학교 컴퓨터공학과
{seattle○, ds2shg, joon, yskim01, cypark}@hallym.ac.kr

A study on The Guarantee of QoS in the Home Network using Multiple Speech
Ji-Soo Hwang○ Chang-sub Lee Joon-Seok Park Yu-Seop Kim Chan-Young Park
Dept. of Computer Engineering, Hallym University

요 약

휴대전화에서 전달되는 음성데이터들이 전달되는 과정에서 잡음 등의 외부 요인으로 인하여 데이터에 손상이 생기는 문제가 발생한다. 이렇게 전달된 음성데이터가 음성 인식기를 통과하면 비로 음성 인식기를 통과했을 때보다 인식률이 낮아진다. 본 연구에서는 음성인식 알고리즘을 이용하여 홈 네트워크를 제어하는데 있어서 음성 인식률을 향상시키기 위해서 반복적으로 음성 데이터를 입력받아, 이를 유사한 음성데이터로 재생하여 추출된 여러 개의 데이터(text)를 이미 구축한 홈 네트워크에서 관련 사전에 등록된 단어와의 유사성을 검토하여 추출한 결과로 홈 네트워크를 제어하는 방안을 제안한다. 이 결과, 기존의 방법과 비교하여 10% 정도의 인식률을 향상시킬 수 있었다.

1. 서론

홈 네트워크는 가정 내의 PC를 비롯한 정보가전기종부터 유선 또는 무선의 네트워크로 연결하여 주변기기 공유 및 상호작용을 가능하게 하며, 인터넷이나 휴대용 정보 단말기들 이용한 외부 네트워크와의 연동으로 언제 어디서나 자유롭게 가정의 디지털 가전기기를 원격 제어할 수 있는 시스템을 말한다.[1]

지급까지 홈 네트워크 시스템에 제어 명령을 보낼 때는 휴대폰이나 인터넷에서 원격을 보고 여러 절차를 통해 버튼을 늘리 조작하도록 되어 있어서 사용이 불편하고 번거로웠다. 현재의 홈 네트워크 기술은 음성통화를 사용함으로써 유무선 전화를 사용할 수 있는 것이지만 어디서나 이용할 수가 있고, 사용자가 원할 때 누구나 쉽게 이용할 수 있다는 장점이 있다. 그러나 이렇게 음성 통화를 이용해서 홈 네트워크 시스템을 제어하는 경우, 특히 전화기를 이용한 음성인식은 입력되는 음성이 일반 마이크와 다르게 전화기를 통하므로 음각이 많고 인식률도 현저하게 감소하고, 전화선의 기본적인 대역폭 때문에 음성 정보가 상당 부분 위축/보완되어 있다.[2] 이러한 점은 음성인식 분야의 발전과 정보통신의 발달과 더불어 해결해 나갈 문제이다. 이러한 과정에서 음성 정보가 음성 인식기를 거쳐 얻어진 결과(text data)를 시점과 비교하였을 때, 사전에 없는 단어일 경우가 발생할 수 있으며 음성 인식기를 거쳐 나온 결과와 이미 구축된 홈 네트워크 관련 정보에 있는 단어와의 유사성을 비교하고, 여러 번 같은 단어를 입력하여 이동 단말 상에서 사용자가 입력한 음성 정보를 정확하게 인식할 수 있도록 하여 인식률을 높여 홈 네트워크 시스템의 제어를 수행하게 하고자 한다.

2. Home Network System

IT 기술의 발달과 고속성장을 통한 인터넷 보급과 더불어 기업이나 공공기관의 사무실을 중심으로 구축된 네트워크 환경이 가정 내의 디지털 전자기기로 확산되었 다.

홈 네트워크는 정보의 처리, 관리, 전달 및 저장에 있어서 가장 대표적인 구현 방식으로, 사용자의 요구에 따라 정의하는 결과를 제공하고, 각기 다른 기기를 연결하여 통합할 수 있게 해주는 구성요소들의 집합이다. 이는 데이터와 통신의 공유 및 상호작용을 가능하게 하는 2개 이상 장비(노드)의 조합으로 이루어진다.

[그림 1] Home Network System

* 이 논문은 2004년도 한국과학회과학기술학회지 학술발표논문집 Vol. 31, No. 2에 의하여 연구되었음.
3. 응성인식 시스템

기계에 의한 응성인식의 결과가 실제로 여러 분야에 응용하는 사례가 많아 넓어지고 있다. 예를 들어, 목소리를 가지고 특정인을 확인한다면, 국제전화를 할 때 송신자의 목소리를 인식하여 수화자의 목소리로 통화할 때 사용한다면, 가전제품을 동작시킬 때 사용자의 음성명령에 따라 구동시키던단지, 여러 사례를 생각해 볼 수 있다. 하지만, 음성이 발화되는 환경은 보통 매우 엄격하여 (자동차 소음, 균치소의 소음, 또는 여러 음이 섞여거나 반사되는 등) 기계가 정확히 음성을 인식하여 그 결과를 출력하는 것이 매우 어려운 일이다.

따라서, 지금까지 구현된 응성인식의 출력은 어느 정도 오류를 포함하고 있으며, 이를 입력으로 하여 위에 언급한 음용에 사용하기 위해서는 오류를 감소하고 현황에 맞게 이를 수정하는 전략이 필요하다.[5]

응성인식 시스템이 마이크를 통한 것이 아닌 전화기로 음성정보를 입력받을 경우에는 인식률이 감소한다. 한 단계 더 나아가서 유선 전화기와 아닌 무선의 휴대전화를 사용하면 인식률은 더 낮아지게 된다.

기존의 응성인식을 이용한 휴대전화 시스템에서는 음성의 휴대전화 환경을 이용하여 음성에 의해 아이디를 인식할 수 있다.

본 연구에서 제안하는 시스템은 사람들의 목소리를 인식하는 휴대전화로 음성데이터를 입력하여 음성 인식기를 거쳐 얻은 음성(data)을 휴대전화 콘서턴트에 있는 단어과 비교하여 가장 적절한 명령어를 휴대전화 시스템에 연결된 각각의 장치를 제어하여 한다.

휴대전화에는 음성의 Encoder/Decoder 알고리즘으로 QCELP 알고리즘을 사용한다. 휴대전화에서 음성인식은 휴대전화에서 사용하는 하드웨어 사양과 기본적인 음성데이터를 입력하여 휴대전화의 음성 인식기를 거친 후에 결과로부터 인식률을 향상시키는 휴대전화 시스템에 제어하는 방법을 제안하고자 한다.

4. 다중발화 응성인식 시스템

응성데이터가 휴대전화를 이용하여 전달될 때 QCELP 알고리즘을 사용하는데 휴대전화에서는 제한된 사양을 가지고 있기 때문에, 고품질의 음성데이터를 전달하는 데에는 문제가 있다. 이러한 알고리즘과 더불어 휴대전화와 같은 휴대전화 환경에서 음성 정보를 전달할 경우 데이터의 손상이나 외부 및 음성정보가 음성 인식기에 입력될 때 외부의 잡음과 같은 변동이 인식률을 떨어뜨리는 요인이 된다.

그러므로, 본 연구에서는 인식률을 향상시키기 위한 방법으로 이동 단말에서 다중 발화로 입력된 음성 정보를 음성 인식기로 입력하여 나온 결과(text data) 에 유사한 알고리즘을 적용하여 인식된 단어의 음을 보다 정확한 단어로 수정하여 휴대전화 시스템에서 각각의 장치를 제어할 수 있도록 하였다.

[그림 2] 다중발화 응성인식 시스템
본 연구에서 제안하고자 하는 시스템의 과정은 다음과 같다.

(1) 이동 단말(휴대 전화 등)을 이용하여 음성데이터를 다중 발화한다.
(2) 단말기에서 입력받은 음성데이터는 home network system의 음성 인식기를 거쳐 텍스트 데이터로 추출된다.
(3) 추출된 텍스트 데이터를 휴네트워크 콘서턴트에 비교하여 해당하는 단어를 찾는다. 이때 유사한 알고리즘을 적용한다.
(4) 해당하는 단어를 찾지 못하는 경우에 비교하려는 다른 단어의 데이터를 가지고 다시 비교한다.
(5) home device를 제어한다.

5. 실험

실현 환경은 보이스트레크에서 제작한 바이오닉스를 사용하고, 휴네트워크에서 사용되는 단어들은 인터넷으로 등록하였다. 본 논문에서 제안하는 다중발화의 단일 발화하였음 경우의 인식률을 비교하기 위해 10명의 피
형자를 모집하였다. 다중발화의 경우 10개의 단어를 각각 3번씩 발생하며 단일 발화의 경우는 10개의 단어를 1번씩만 발생하여 음성 인식기를 통해 나온 text data를 각각 읽을 수 있었다.

실험 과정에서 발생된 음성 정보는 음성 인식기를 거쳐 text data로 변환된다. 읽어진 text data는 유사율 알고리즘을 적용하지 않아도 어느 정도의 인식률을 확인할 수 있었다.

그러나, 인식률을 높이기 위하여 읽어진 단어(text data)는 데이터베이스를 기반으로 음성 인식기의 성능을 향상시키기 위해 시스템의 향 내트워크 시스템 제어를 위해 관련 음성 사전과의 유사율을 검사한다. 유사율의 threshold 값을 50%로 설정하였다. threshold 값이 50% 이상이면, 오인 식 단어로 판정하고 음성 사전에 있는 단어로 수정하여 처리하는 경우가 포함된다.

유사율은 다음과 같이 구할 수 있다.[4]

\[
\text{유사율(\%)} = \frac{\text{일치율결수}}{\text{주체단단의 전체음절수}} \times 100
\]

[그림 3]은 text data에 유사율 알고리즘을 적용했을 경우와 적용하지 않았을 경우의 인식률의 차이를 보여주고 있다.

[그림 3] 유사율 알고리즘 적용 유무에 따른 비교


[그림 4]에서 보는 바와 같이, 유사율 알고리즘을 적용한 후에도 1회 발생 후의 인식률은 89.75%로, 3회 발생 후의 인식률은 93%로, 10번의 인식률의 향상을 볼 수 있다. [그림 4]에서와 같이 형광등, 오디오, 키보드 경우에는 1번 발화보다는 3번 발화함으로써 인식률을 높일 수 있었다.

[그림 4] 1회 발성 인식률과 3회 발성인식률의 비교

6. 결 론

전화와 같은 유선의 네트워크 환경에 비해서 유선에서 음성 데이터를 입력할 때 데이터의 손실이나 왜곡의 정도가 크다. 그리고, 흡박크릭에서 사용하는 고전적인 인식 알고리즘에서 나타 นอกจาก나 제약을 고려하여, 본 연구에서는 사용자가 이용할 때 다중발화 음성 정보를 음성 인식기에 입력시키며 나온 결과(text data)를 구축되어 있는 흡박크릭 음성 사전과 비교하여 유사율 알고리즘을 적용 후의 인식률은 단일발화 했을 경우에 비해 10%정도의 향상이 있었다.

하지만, 사용자의 측면의 QoS를 고려해 볼 경우 여러 번 같은 단어를 입력해야 한다는 문제점이 있어서 추후에는 사용자가 여러 번 같은 단어를 입력하는 방법이 아닌 흡박크릭과 같은 이동 단말에서의 입력방법 음성데이터를 더 효율적으로 인식할 수 있는 방법을 연구하고자 한다.

7. 참고문헌