

## 音聲工學의 展望

安秀桔

서울대학교 대학원 전자공학과

### Prospects of Speech Processing

SouGuil ANN

Dept. of Electronics Eng., Seoul National University

#### I. 서론

音聲工學의 미래를 예측한다는 것은 생각보다 광범위하고 막대한 것이 된다. 앞으로 음성공학의 발달이 주는 영향은 막대해서 그 미래를 이야기한다는 것은 마치 그레함벨 이전에 전화의 도래를 이야기하는 만큼 획기적인 것이 될 것이다. 그러나 현실적으로 얻어지는 결과는 누가 어떠한 필요성에 의해서 어느 만큼 투자를 하느냐에 따라 음성공학발달의 영향의 전폭적인 도래시기에는 미지요소가 많이 개입하게 될 것이다. 어느 나라가 어느 만큼 빨리 구현을 시키느냐 하는 시기에는 차가 있으나 가는 길에 대강 같게 된다.

#### II. 음성공학의 필요성과 전망

音聲工學研究의 필요성은 크다. 그것은 근래 computer가 많이 발달해서 인간 가까이에서 여러가지 써-바스를 해 주고 있는 데 이 때 computer가 인간에게 어떠한 통보를 하고자할 때 지금까지 해온 것과 같이 화면을 통해서 행할 경우에는 사람의 시선을 reading 한가지에 독점해 버리는 폐단이 있기 때문이다. 사람에게 가장 편한 형태는 음성에 의한 입출력이다.

단 여러-가 많이 생기기 때문에 다시말해 人間이 잘 못 알아듣고 잘 못 짐작하기때문에 문제가 있게 되는 데 그러한 경우 대화가 interactive해서 틀린 것이 수정된다는 것이 가장 이상적이 될 것이다.

이때에는 computer로 부터 음성이나 기타 형태로 인간에게 의사전달 또는 결정전달을 했을 때 사람이 복창을 해서 기계가 그것을 판독해야하기 때문에 기계의 음성인식기능이 요구된다.

기계가 인간에게 통신을 위해서 음성을 쓸 경우에는 음성합성기술이 필요하다. 그것은 제법 해결되어 있는 사례가 있다. 우리나라에서도 몇가지 제품이 나오고 있다 [1]-[5]. 그러나 일반적으로 합성된 음성을 알아 듣기 위해서는 비교적 짧으나마 청취자의 적응기간이 필요하게 된다.

처음부터 100% 알아듣게 되지 않아서 기계사투리에 적응해야 한다고 보면 될 것이다. 한국어의 경우에도 음질개선은 충분히 가능하지만 이를 위해서 투자를 결심시키기에는 시장규모가 문제가 된다. 발음의 질 때문에 문제가 되는 것은 결국은 사람의 재훈련으로 낙착이 되기 때문에 사용을 거부하는 사례가 생기게 된다. 편리하니까 쓰지만 그렇게 되면 결국은 사람이 기계사투리에 따라가게 된다는 폐단이 있게 되는것이다. 사람의 적응능력이 크기 때문에 기계의 열악발음시대가 길면 사람발음에 영향을 주게 될 우려도 있다.

현 시점에서 사람언어의 전폭적인 인식은 요원하다. 그러나 극히 한정된 인식은 어느 정도 가능하다 [6]-[10]. 문제는 몇 단어를 인식을 시킬 것인가 하는 것이 문제가 될 따름이다.

그 다음에는 어떠한 정도의 인간의 협조를 기대할 수 있는나에 따라서 달라진다. 여러번 발음을 되

풀이 시킬 수 있는 조건인지 아니면 한 마디로써 알아 들어야 하는 것인지에 따라 다르다. 사람말을 알아 듣는 로보트도 써-비스 인원과 같이 써-비스의 질이 문제가 되고 이때 사람의 요구는 가혹할 정도 까다롭게 되어 사람에게 요구하는 정도보다 더 낮가를 요구 받게된다. 그것은 시장의 원리에 의해서이다.

또 하나의 형태는 적극적으로 인간이 computer에 맞춰주는 형태이다. 경우에 따라서는 제일 먼저 도달하는 형태가 될 가능성이 있는 데 그것은 외국에서 개발되어 대량생산을 통해서 값싸게 생산된 것을 쓰게 되는 경우로서 그 경우에는 사용자가 로보트를 위해서 영어를 배우게 되는 경우이다.

한국어의 경우도 미완성의 기계가 상업적으로 성공을 거두면 사람들이 기계가 인식하기 쉬운 발음으로 사람이 적응하게 되어서 기계조작을 좋아하는 사람 또는 computer광이 되어 사람보다는 기계를 상대하는 시간이 많은 사람들은 발음이 달라지고 기계가 쉽게 알아듣는 발음으로 변천되 가게 된다.

사용자가 자기의 발음을 적응시켜 computer를 기막히게 잘 쓰게 되는 사람이 생기기 전에 사람의 표준어로 작동되는 기계를 만드는 것은 산업이 어느 만큼의 시장으로 파악하느냐에 또는 정부가 한국어의 identity 수호를 어느 정도의 중요과제로 생각하게 되느냐에 따라서 달라진다.

전술한바와 같이 시장으로서는 큰 것이 아니기 때문에 그러한 자본과 기술을 갖고 있는 한국회사라면 기업으로서는 영어시장을 겨누는 것이 건전하다고 보아야 할 것이다. 영어의 국제어화 경향에 따라 영어 시장은 무척 큰 것이 된다.

정부의 부자의 경우는 앞으로 국민의 압력이 강화되는 마당에서 결 공은(고품질 써-비스) 음성인식에 필요한 막대한 예산을 돌린다는 것은 어려워 보인다. 그리고 그것에 성공할 정도의 압력구축을 위한 인원이 있어 보이지 않는다.

결국은 한국어를 위한 기계-인간간 음성 interface는 영어의 경우에 비해서 무척 처질 것으로 보인다. 통신에 있어서 ISDN의 구현이 처진 것과 마찬가지로 다른 나라에 비해서 늦은 완성을 보게 되겠

지만 기술실력의 부족만이 아니고 소비자에게 어떤 정도 어필이 되어서 소비자들에게 돈을 내게 하느냐에 따라서 결정되기 때문이다.

이러한 여건하에서 늦게 도래할 또는 가까운 단계에서는 이룩되지 않을 가능성도 있는 음성 interface기술의 목표로서는 다음과 같은 것이 있을 수 있다.

1. 신문, 소설, 시 낭독: 원하는 음성과 음색으로 신문이나, 소설 또는 시를 낭독케 함.
2. Scheduler, Checker: 당일 또는 향후의 schedule를 받아 기록하고 remind를 시킨다. 완성분에 대해서 음성으로 다시 등록하고 미완성부분을 check한다.
3. 기기작동: 가정에서 말에 의해서 TV채널을 바꾸고 손에 짐이 있을 때에 문을 열고 불을 켜고 수도물을 흘리며 오븐의 스위치를 넣는다. 목욕탕물을 대우기 시작한다.
4. 원격작동: 전화를 통해서 가정의 기기를 미리 작동시킨다. 자동차의 엔진을 미리 걸게 원격시달을 한다.
5. 은행거래잔액조회와 음성보증: 원격리에서 잔고를 확인하고 은행수표 발행시 그 한도를 전화를 통해 알린다. 또는 음성을 통해 거래 확인도장을 찍는다.
6. 번역전화: 화자의 목소리로 그리고 화자의 감정기복을 그대로 반영하는 목소리로 외국어로 번역해서 말 해주는 장치, 이경우에 비영어의 경우 정부간의 특별고려가 없는 한 번역대상은 영어로 한정 되겠다.
7. 초감정 의사전달: 원하는 어조로 의사전달을 한다.
8. 음성호출전화: 호출자에게 가장 편한 호칭으로 들먹이는 것으로 전화를 해준다. 복창을 시켜 맞을 때에 확인을 해주면 dial을 시작한다.
9. 음성타자기: 구술함에 따라 타자를 찍어 준다. 역시 복창을 시켜 틀린 데가 있으면 시정하다. 단 이경우에는 대형computer가 필요 하다. 화면표시를 통해서 교정하는 경우는 교정기능부분만이라면 소형계산기도 가능하다.

10. 음성명령시달: 자동차 등을 운전하면서 보조기능을 작동시킨다.
11. 음성교정: 발음훈련과 발음교정을 해준다. 연설에서 감정전달을 시정해준다.
12. 감정판정기능: 상대편의 감춰진 감정의 기조를 알게 한다.
13. 언어력판정: 화자의 출신과 성장력을 찾아 낸다.
14. 음성을 통한 작업보조: 음성을 통해서 요구된 계반작업을 도와준다. 작업 경과를 보고한다. 자료의 display를 음성 명령에 따라서 한다.
15. 음성을 통한 발성기관, 또는 일반신체조건을 체크하고 진단을 한다.

### III. 결론

음성이 사람에게 있어서 가장 자연스러운 의사교환방법임을 감안 해서 음성처리기술의 발달은 사람들에게 가장 편한 자세와 눈 감고 편한 방법으로 다른 일을 하면서 겸해서 일을 할 수 있게 해주고 보지 않는 사용자에게도 음성으로 주의를 환기시켜 시정과 개입을 하고 개입을 허용하는 점으로 해서 가장 이상적이다.

그러나 각각의 단계에서 상당한 투자가 드는 문제로 해서 상기의 모든 항목이 구현되는 가능성에는 차이가 있다.

### 참고문헌

- [1] 이정석 외, "한국어 text-to-speech 합성기의 구현," 한국정보과학회 춘계학술발표논문집, vol. 19, pp. 107-110, 1992.
- [2] 안승권, 성경모, "포맷트 궤적 중첩방법에 의한 한국어 문자-음성 변환," 제9회 음성통신 및 신호처리 워크샵 논문집, 1992.
- [3] 최운현 외, "고품질의 한국어 문장음성변환 시스템: 글소리II," 제9회 음성통신 및 신호처리 워크샵 논문집, 1992.

- [4] 구준모 외, "한국어 무제한 음성합성 시스템: 가라사대," 제9회 음성통신 및 신호처리 워크샵논문집, 1992.
- [5] 조철우 외, "A study on the implementation of Korean synthesis-by-rule system using formant synthesis method," 한국음향학회지, vol. 9, no. 6, 1990.
- [6] 황영수, "TDNN과 DTW를 이용한 격리단어 인식," 한국음향학회지, vol. 12, no. 2, 1993.
- [7] 김남수, 은종관, "상호정보를 이용한 Hidden Markov Model 파라미터의 트레이닝과 음성인식에의 적용," 한국음향학회지, vol. 9, no. 5, 1990.
- [8] 안태욱, 이상훈, 김순협, "VQ와 multi-layer perceptron을 이용한 단모음 인식에 관한 연구," 한국음향학회지, vol. 12, no. 1, 1993.
- [9] 이가영, 최갑석, "뉴럴-퍼지패턴매칭에 의한 단어 인식에 관한 연구," 전자공학회 논문지, vol. 29-B, no. 11, 1992.
- [10] 김석동, 이행세, "복합신경망을 이용한 단음인식에 대한 연구," 한국음향학회지, vol. 11, no. 6, 1992.