음성공학의 전망

안 수 길 (서울대 전자공학과)

음성공학의 미래를 예측한다는 것은 생각보다 광범위한 것이 된다. 사람이 가장 편하게 자기의 의사를 전달하는 형태가 상대를 보지도 않고 말로서 일을 시키는 것인 데 이러한 방법으로 computer를 활용하게 되는 것은 획기적인 뜻이 있다.

그것은 백년전에 전화의 도래를 이야기하는 만큼 중요한 것이다. 그것은 누구가 어떠한 필요성에 의해서 어느 만큼 투자를 하느냐에 따라 도달시기에는 미지요소가 많이 개입하기 때문에 실제 발달의 속도는 말하기가 어렵다.

필요성은 크다. 근래 많이 computer가 발달해서 인간가까이에서 여러 가지 써비스를 해주고 있지만 그 computer가 인간에게 어떠한 통보를 하고자할 때 지금까지 해온 것과 같이 화면을 통해서 행할 경우에는 사람의 시선을 reading 한가지에 독점해 버리는 폐단이 있기 때문이다. 즉 사람이 가장 편한 형태는 음성에 의한 입출력이다. 또는 같은 내용도 음성pitch를 통해서 더 감정을 담고 또는 위협적으로 전달되는 것을 선호하는 경우도 있게 된다.

단 에러가 많이 생기기 때문에, 다시 말해 인간이 잘못 알아듣고 잘못 짐작하기 때문에 문제가 있게 되는데 그러한 경우 대화가 interactive해서 틀린 것이 수정된다는 것이 가장 이상적이 될 것이다. 이 때에는 computer로부터 음성이나 기타 형태로 인간에게 의사전달 또는 결정전달을 했을 때 사람이 복창을 해서 기계가 그것을 판독해야 하기 때문에 기계의 음성인식기능이 요구된다.

기계가 인간에게 통신을 위해서 음성을 쓸 경우에는 음성합성기술이 필요하다. 그것은 제법 해결되어 있는 사례가 있다. 우리나라에서도 몇 가지 제품이 나오고 있다. 그러나 일반적으로 합성된 음성을 알아 듣기 위해서는 비교적 짧으나마 청취자의 적응기간이 요구된다. 기계가 한 말을 알기란 어려워서 처음부터 100% 알아듣게 되지 않아서 일종의 기계사투리에 적응해야 한다고 보면 될 것이다.

음질개선은 충분히 가능하지만 이를 위해서 투자를 하려면 시장규모가 문제가 된다. 발음의 질 때문에 문제가 되는 것은 결국은 사람의 재훈련으로 낙착이 되기 때문에 자연감이 충분치 못 하면 사람들이 그러한 기계의 사용을 거부하는 사례가 생기게 된다. 사람의 적응능력이 크기 때문에 기계의 열악발음시대가 길면 사람발음에 영향을 주게 될 우려도 있다.

현 시점에서 사람언어의 전폭적인 인식은 요원하다. 그러나 극히 한정된 어휘의 인식은

어느 정도 가능하다. 문제는 몇 단어를 인식을 시킬 것인가 하는 것이다.

그 다음에는 어떠한 정도의 인간의 협조를 기대할 수 있느냐에 따라서 달라진다. 여러번 발음을 되풀이 시킬 수 있는 조건인지 아니면 한 마디로써 알아들어야 하는 것인지에 따라 다르다. 어려운 사람에게 기계시종을 시킬 수는 없다.

사람말을 알아듣는 로보트도 써비스 인원과 같이 써비스의 질이 문제가 된다. 한 마디로 간단히 알아들으면 좋지만 그렇지 못 할 경우는 사용을 기피하게 된다.

또 하나의 형태는 아쉬운 사람들이 쓰는 경우인데 기계가 발성할 경우나, 반대로 사람이 말하는 것을 기계가 인식할 경우나 적극적으로 인간이 computer에 맞춰주는 형태이다. 경우에 따라서는 제일 먼저 도달하는 형태가 될 가능성이 있는 데 그것은 외국에서 개발되어 대량생산을 통해서 값싸게 생산된 것을 도입해서 쓰게 되는 경우로서 그 경우에는 사용자가 로보트를 위해서 영어를 배우고 또 로보트가 알아 듣는 accent로 애써 발음을 맞춰주게 되는 경우이다.

한국어의 경우도 미완성의 기계가 상업적으로 성공을 거두면 기계가 인식하기 쉬운 발음으로 사람이 적응하게 되기가 쉽다. 기계조작을 좋아하는 사람 또는 computer광이 되어 인간보다는 기계를 상대하는 시간이 많은 사람들은 발음이 달라지고 기계발음에 따라가거나 기계가 쉽게 알아듣는 발음으로 변천되어 가게 된다.

자기의 발음을 필요한 대로 변화시켜 computer를 기막히게 잘 쓰게 되는 사람이 생기기 전에 사람의 표준어로 작동되는 기계를 만들어야 하는 데 그것은 산업이 어느 만큼의 시장으로 보고 투자를 결심하느냐, 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 채산이 달아서 기업으로는 건전한 것이다.

정부의 투자의 경우는 앞으로 정부의 예산낭비 여부에 국민의 압력이 강화되는 마당에서 고품질 써비스 음성인식에 필요한 막대한 예산을 돌린다는 것은 기대하기가 어렵다. 그리고 그러한 행정결정에 성공할 정도의 압력구축을 위한 관심인구가 있어 보이지 않고 예산투입을 결심해도 충분한 연구자인원수가 양성되어 있지 못하다.

결국은 한국어를 위한 기계-인간간 음성interface는 영어의 경우에 비해서 시간적으로 무척 처질 것으로 보인다.

통신에 있어서 ISDN의 처지도 마찬가지로 다른 나라에 비해서 늦은 완성을 보게 되겠지만 기술실력의 부족만이 아니고 소비자에게 어떤 정도 appeal이 되어서 소비자들에게 돈을 내게 하느냐에 따라서 결정되기 때문이다. 말하자면 보급률이 어느 만큼 빨리 올라갈 수 있느냐하는 문제이다.

이러한 여건하에서 늦게 도래할, 또는 끝내 이룩되지 않을 가능성도 있는 음성interface 기술의 목표물로서는 다음과 같은 것이 있을 수 있다.

미래의 응용예

1. 신문.소설.시 낭독: 원하는 음성. 음색으로 신문이나, 소설 또는 시를 낭독케 함.

2. Scheduler, Checker: 당일 또는 향후의 Schedule를 받아 기록하고 remind를 시킨다. 완성 또는 수행분에 대해서 음성으로 다시 등록하고 미완성부분을 check 또는 remind시킨다.

3. 기기작동: 가정에서 말에 의해서 TV채널을 바꾸고 손에 짐이 있어 손이 available 하지 않을 때에 문을 열고 불을 켜고 수도물을 흘리며 오븐의 스위치를 넣는다. 목욕탕물을 미리 데우게 한다.

4. 원격작동: 전화를 통해서 가정의 기기를 미리 작동시킨다. 겨울에 자동차의 엔진을 시동시켜 엔진을 미리 데우게 원격시달을 한다.

5. 은행거래잔액조회와 음성보증(confirmation): 원거리에서 잔고를 확인하고 은행수표 발행시 그 액수 또는 한도를 전화를 통해 알려 놓은 다. 또는 음성을 통해 거래 확인도장을 찍는 다.

6. 번역전화: 화자의 목소리로 그리고 화자의 감정기복을 그대로 반영하는 목소리로 외국어로 번역해서 말해 주는 장치, 이경우에 정부간의 특별고려가 없는 한 번역대상은 우선 영어로 될 것이다.

7. 초감정 의사전달: 감정이 격할 때에도 부드러운 소리로 또는 자기가 원하는 어조로 의사전달을 한다. 적시에 종종 "자기" 또는 "선생님"이라고 말 하면 기계가 상대편의 이름을 그 때마다 넣어 주어 친밀감을 더해 준다.

8. 음성호출전화: 호출자가 자기에게 가장 편한 호칭 또는 별명으로 피호출자의 이름을 말 하는 것으로 전화를 dial해 준다. 복창을 시켜 맞일 때에 확인을 해 주면 dial을 시작해서 전화접속을 해 준다.

9. 음성타자기: 사람이 말 하는 것에 따라 타자를 찍어 준다. 역시 복창을 시켜 틀린 데가 있으면 시정한다. 단 이 경우에는 대형computer가 필요하다. 화면표시를 통해서 교정하는 경우는 교정기능부분만이라면 소형계산기도 가능하다.

10. 음성명령시달: 자동차 등을 운전하면서 보조기능을 말로 요청하면 기계가 알아듣고 그 기능을 작동시킨다.

11. 음성교정: 발음훈련과 발음교정을 해 준다. 연설에서 감정전달을 시정해 준다.

12. 감정판정기능: 상대편의 감춰진 감정의 기조를 알게 한다.

13. 언어력판정: 화자의 출신지방, 가정의 정도와 성장력을 찾아 낸다.

14. 음성을 통한 작업보조: 음성을 통해서 요구된 제반작업을 도와 준다. 작업경과를 보고한다. 자료의 display를 음성명령에 따라서 한다.

15. 음성을 통한 발성기관, 또는 일반신체조건을 검사하고 진단을 한다.

검 토

음성이 사람에게 있어서 가장 자연스러운 의사교환방법임을 감안할 때 음성처리기술의 발달은 사람들에게 가장 편한 자세와 눈 감고 편한 방법으로 다른 일을 하면서 겸해서 일을 할 수 있게 해 주고 보고 있지 않는 사용자에게도 음성으로 주의를 환기시켜 시정과 개입을 하고 또는 자기편에서도 개입을 허용하는 점으로 해서 가장 이상적이다.

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

결 론

Computer발달의 최종단계에서는 음성에 관한 모든 양상이 기계에서 파악되어야 한다. 음성을 통한 사람들의 의사소통은 상대편의 생각 즉 사고방식에 대해서 상당한 정보를 또는 선행공유지식을 청취자가 갖고 있어서 들을 때에 많은 추리가 선행되어야 하고 들리는 말만 완전하게 파악하는 것으로는 불충분하다. 즉 아무리 정확하게 알아듣는 청취자도 (기능만을 획득하고 어휘만 충분한 어린이의 경우와 같이) 일반사람, 또는 화자에 관한 상식이 없어서는 이해가 완벽하지 못 하다.

따라서 말의 내용의 파악을 위해서는 상당한 용량의 computer를 갖고 있어서 이를 미

리 일반인 정도는 길러놓아야 한다는 뜻이다.(learning process).

이 speech recognizer의 도움을 받아서 편하게 활용할 수 있기 위해서는 상당한 투자가 필요한 데 일반상품화를 위해서는 과도한 액수가 된다. 아직 특수기관등에서 고액의 투자를 하고 그 기계가 상품으로 완성된 다음 그 knowhow가 업자의 지출비를 현저하게 줄여주기 전에는 가능한 날자가 무척 뒤로 처지게 된다.

기계사투리를 상당 정도 허용한다면 음성합성기는 비교적 단시일내에 활용이 되고 나아가서는 음질의 개선도 기대할 수도 있을 것이다.

음성공학의 본격적인 도움을 얻을 수 있게 된다면 그것은 인간의 지금까지의 한계를 확장해서 음성의 가공, 또는 심도 있는 분석을 통해서 인간의 exploitation에 가까운 양상이 드러나 도덕적 문제가 제기될 것이다. 다시말해 발달만을 바랄 수가 없는 양상이 있다.

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	N		
Trill	B			r					R		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	ʼ	as in:
|	Dental	ɗ	Dental/alveolar	pʼ	Bilabial
!	(Post)alveolar	ʄ	Palatal	tʼ	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	kʼ	Velar
‖	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

SUPRASEGMENTALS

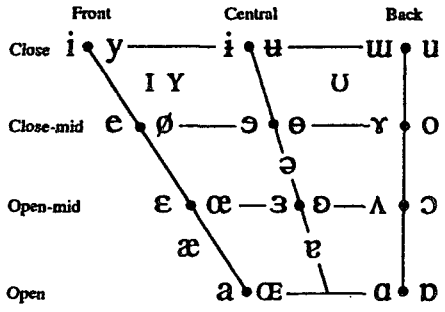
ˈ	Primary stress	ˌfounəˈtɪʃən
ˌ	Secondary stress	
ː	Long	eː
ˑ	Half-long	eˑ
◡	Extra-short	ĕ
.	Syllable break	ɹi.ækt
|	Minor (foot) group	
‖	Major (intonation) group	
‿	Linking (absence of a break)	

TONES & WORD ACCENTS

LEVEL		CONTOUR	
é or ˥	Extra high	ě or ˩˥	Rising
é ˦	High	ê ˥˩	Falling
ē ˧	Mid	ě ˦˥	High rising
è ˨	Low	ȅ ˩˨	Low rising
ȅ ˩	Extra low	ȇ ˥˦˥	Rising-falling
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

etc.

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

OTHER SYMBOLS

ʍ	Voiceless labial-velar fricative	ɕ ʑ	Alveolo-palatal fricatives
w	Voiced labial-velar approximant	ɺ	Alveolar lateral flap
ɥ	Voiced labial-palatal approximant	ʄ	Simultaneous ʃ and x
ʜ	Voiceless epiglottal fricative		
ʢ	Voiced epiglottal fricative		
ʡ	Epiglottal plosive		

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.

k͡p t͡s

DIACRITICS

Diacritics may be placed above a symbol with a descender, e.g. ŋ̊

◌̥	Voiceless	n̥ d̥	◌̤	Breathy voiced	b̤ a̤	◌̪	Dental	t̪ d̪
◌̬	Voiced	s̬ t̬	◌̰	Creaky voiced	b̰ a̰	◌̺	Apical	t̺ d̺
◌ʰ	Aspirated	tʰ dʰ	◌̼	Linguolabial	t̼ d̼	◌̻	Laminal	t̻ d̻
◌̹	More rounded	ɔ̹	◌ʷ	Labialized	tʷ dʷ	◌̃	Nasalized	ẽ
◌̜	Less rounded	ɔ̜	◌ʲ	Palatalized	tʲ dʲ	◌ⁿ	Nasal release	dⁿ
◌̟	Advanced	u̟	◌ˠ	Velarized	tˠ dˠ	◌ˡ	Lateral release	dˡ
◌̠	Retracted	i̠	◌ˤ	Pharyngealized	tˤ dˤ	◌̚	No audible release	d̚
◌̈	Centralized	ë	◌̴	Velarized or pharyngealized	ɫ			
◌̽	Mid-centralized	e̽	◌̝	Raised	ę (= voiced alveolar fricative)			
◌̩	Syllabic	n̩	◌̞	Lowered	ę (= voiced bilabial approximant)			
◌̯	Non-syllabic	e̯	◌̘	Advanced Tongue Root	e̘			
◌˞	Rhoticity	ə˞	◌̙	Retracted Tongue Root	e̙			

한 글 음 성 문 자

KOREAN PHONETIC ALPHABET

	방법 \ 위치	두 입술	입술·이	이/잇몸	혀말음	뒤 잇몸	잇몸입천장	센입천장	여린입천장	목 청	목구멍	소리문
닿소리	터짐소리	ㅂ ㅃ ㅍ ㅸ		ㄷ ㄸ ㅌ ㅶ	ㄷ ㄸ ㅌ ㅶ			ㄱ ㄲ ㅋ ㅷ	ㄱ ㄲ ㅋ ㅷ	ㄱ ㄲ ㅎ ㆅ		ㅇ
	터갈이소리					ㅈ ㅉ ㅊ ㅉ						
	갈이소리	ㅱ ㅱ	ㅿ ㅿ	ㅅ ㅿ ㅆ ㆍ ㆍ	ㅅ ㅿ ㅆ	ㅊ ㅊ	ㅅ ㅿ ㅆ	ㄱ ㄱ	ㄱ ㄱ	ㅎ ㅎ	ㅧ ㅧ	ㆆ ㆆ
	튀김소리			ㄹ	ㄷ					ㄹ		
	굴림소리			ㄹ						ㄹ		
	혀옆소리			ㄹ	ㄷ			ㄹ				
	콧소리	ㅁ	ㅇ	ㄴ	ㄴ			ㄴ	ㅇ	ㅇ		
	지속음/반홀소리	ㅜ ㅗ	ㅡ	ㆍ	ㄷ			ㅣ	ㅜ	ㅎ		

홀소리	\			
	닫힌	ㅟ ㅠ ㅜ	ㅣ ㅟ ㅠ ㅜ	
	반닫힌	ㅚ ㅗ	ㅔ ㅚ ㅓ ㅗ	
	반열린	ㅙ ㅛ	ㅐ ㅙ ㅝ ㅛ	
	열린	ㅙ ㅛ	ㅏ ㅙ ㅑ ㅛ	

앞 가운 뒤