

한국어 TTS 시스템의 객관적인 성능평가를 위한 기초검토

고락환* 김영일** 김봉완** 이용주***
* 원광대학교 컴퓨터공학과 휴먼인터페이스연구소
** 원광대학교 음성정보기술산업지원센터
*** 원광대학교 전기전자 및 정보공학부

Basic consideration for assessment of Korean TTS system

LagHwan Ko*, YoungIl Kim**, BongWan Kim**, Yong-Ju Lee***

* Human Interface Lab. Department of Computer Engineering, Wonkwang Univ.

** Speech Information Technology & Industry Promotion Center

*** Department of Electronic and Information Engineering, Wonkwang Univ.

* silpyd@nate.com **yikim@sitec.or.kr, bwkim@sitec.or.kr, ***yjlee@wonkwang.ac.kr

Abstract

Recently due to the rapid development of speech synthesis based on the corpora, the performance of TTS systems, which convert text into speech through synthesis, has been enhanced, and they are applied in various fields. However, the procedure for objective assessment of the performance of systems is not well established in Korea.

The establishment of the procedure for objective assessment of the performance of systems is essential for the assessment of development systems for the developers and as the standard for choosing the suitable system for the users.

In this paper we will report on the results of the basic research for the establishment of the systematic standard for the procedure of objective assessment of the performance of Korean TTS systems with reference to the various attempts for this project in Korea and other countries.

I. 서론

음성에 의한 정보전달은 사람간의 커뮤니케이션 방법 중 가장 기본적이며 친숙한 수단이다.

특히 이러한 음성을 인간과 기계간의 인터페이스 수단으로 사용하기 위한 음성인식 및 합성의 연구는 아직도 해결해야 할 많은 과제를 안고 있기는 하지만 응용에 따라서는 실용화나 상품화가 활발하게 이루어지고 있는 것도 있다. 특히 음성합성의 경우에는 임의의 텍스트를 음성으로 변환하는 TTS(Text to Speech conversion) 시스템의 경우, 대량의 음성코퍼스를 기반으로 파형을 접합하는 방식이 가능해짐에 따라 성능이 대폭 개선되어 여러 분야에서 상품화가 활발하게 이루어지고 있다.

그러나 이러한 시스템을 이용할 경우에 여러 제품 중에서 목적에 맞는 것을 선택하기 위한 지침으로서 그 성능을 객관적이며 정량적으로 평가하는 것은 중요한 문제이다. 이는 최종 소비자의 입장에서 뿐만 아니고 응용시스템을 개발하는 사람들의 입장에서도 시스템의 최적한 설계를 위해서 필요하다. 또한 TTS 시스템 자체를 연구 개발하는 입장에서 볼 때 시스템의 개선을 위한 진단적인 평가의 면에서도 성능의 평가는 중요하다.

일반적으로 TTS 시스템은 입력된 임의의 텍스트의 언어적 정보를 해석하여 적절한 휴지구간을 선정하여야 하며, 각종 기호나 숫자들을 포함하여 소리 나는

대로 읽기 형태로 변환하여야 한다. 이를 바탕으로 해당하는 발성내용 및 적절한 운율정보를 포함하는 음성 조각을 미리 저장된 음성 코퍼스로부터 찾아서 이를 접합하는 과정을 통하여 입의 음성을 합성해 낸다. 본고에서는 TTS방식의 합성시스템의 평가 기준 마련을 위하여 각 부분 별로 평가 대상 및 항목들을 설계하는데 고려하여야 할 점들을 살펴보고자한다.

II. TTS의 각 기능별 평가의 검토

1. 텍스트 해석의 평가

일반적인 TTS시스템에서는 입력된 텍스트를 언어적으로 해석하여 읽기규칙 및 기본적인 운율정보의 추정 에 필요한 정보를 얻어서 음성을 합성한다.

텍스트로부터 음성을 합성하였을 때, 그 음성이 알아듣기 어려울 경우, 텍스트 해석에 에러가 있을 경우와 그 이후의 합성단계에서 에러가 있을 경우로 나눌 수 있다. 또 합성음성이 부자연스럽게 들릴 경우도 엑센트나 장단음 등 운율정보의 추정에 에러가 있을 수 있는 등 텍스트 해석이 문제가 될 수 있다. 따라서 텍스트 해석의 평가에서는 제대로 읽기 위한 정보가 올바르게 얻어졌는가의 관점에서 평가하는 것을 목표로 하여야 할 것이다.

TTS 시스템의 응용은 텍스트의 내용을 보지 않고 음성만 듣고 내용을 이해하는 것을 목적으로 하는 경우가 대부분이므로 특히 읽기부분에서 에러가 생기면 텍스트 내용을 이해할 수 없게 된다. 따라서 정독율은 이해성 및 자연성 이상으로 중요한 문제이다.

텍스트 해석의 평가에서는 다음과 같은 점을 검토할 필요가 있다.

- 평가대상이 되는 텍스트의 타입을 한정할 것인가
- 평가는 어떤 관점에서 어떤 항목으로 할 것인가
- 평가는 어떻게 정량화 할 것인가.

1) 평가대상으로 할 텍스트

표나 그림 안의 내용을 읽어주는 형태는 또 다른 고도의 언어 처리가 별도로 연구되어야 하므로 현시점에서는 단순히 문자열로만 구성된 것을 1차 대상으로 하여야 할 것이다.

2) 평가관점 및 평가항목

텍스트 해석에 기인하는 에러의 평가법에는 사용자의 관점과 개발자의 관점이 있을 수 있다. 사용자 관점에서의 평가는 에러의 현상을 표층적으로 평가하는 것으로서, 평가항목으로는 다음과 같은 것이 있을 수 있다.

- 읽기에러
- 엑센트 에러(특히 우리말의 경우 장단음)
- 포우즈 에러

개발자 관점에서는 에러 현상의 평가만이 아니라 그 원인을 분석하여 텍스트 해석에 포함된 각 처리단계의 정확도를 평가할 필요가 있고 이는 성능 개선의 기초 데이터로 사용될 수 있다.

텍스트해석에 따른 에러의 원인을 분류하면

- 사전의 불충분(미등록, 등록에러 등)
- 형태소 해석에러
- 구문해석 에러
- 의미처리 에러

로 대별될 수 있을 것이다.

이들 원인의 분류는 상호 연관되어 있는 경우가 많아서 명확히 구분하는 것은 쉽지 않은 문제이다. 따라서 에러원인을 바탕으로 하는 평가는 평가자의 지식과 판단에 의존하기 쉬우므로 객관적인 평가로서 다루기는 어렵다고 생각된다. 따라서 우선은 에러 현상의 표층적 평가를 대상으로 하는 것이 좋을 것으로 보인다.

3) 평가의 정량화

사용자 관점의 평가에서 중요한, 읽기 에러, 엑센트 에러, 포우즈 에러에 관한 평가의 정량화에 관하여 살펴본다.

읽기에러에서 오독률(잘못 읽은 비율)은 텍스트문장의 크기 당 오독의 수로 말할 수 있지만 우리말의 경우 오독수를 어떤 단위로 할까를 검토해야 할 것이다

가능한 단위로는 음절, 형태소, 어절, 문장 또는 단어에 상관없이 에러갯수 등을 생각할 수 있다.

엑센트의 경우도 같은 문제가 있으나 에러 원인되는 갯수를 대상으로 하는 것이 타당할 것으로 보인다.

포우즈 에러의 평가는 있어야 할 포즈가 없는 경우와 부자연스런 포우즈의 삽입을 모두 고려하여야 하나 이를 정량화하기 위해서는 “있어야 할 포즈의 탈락”과 “부자연스런 포우즈의 삽입”에 대한 판단기준을 설정하기가 쉽지 않다.

2. 이해성의 평가

음성합성의 목적은 음성에 의해 정보를 전달하는 것이므로 정확한 내용이 전달되는가를 평가하는 것은 가장 중요한 사항이다. 일반적으로 이해성의 평가에서 무의미한 음절을 단위로 하여 시험음성을 들려주고 올바르게 들은 음절의 비율을 명료도라고 부르며 의미 있는 단어나 문장으로 구성된 시험음성을 대상으로 한 경우 이해도라고 부르고 있다.

1) 음절명료도

음절명료도의 평가에는 평가자, 시험법, 표시법, 평가 조건 등을 중심으로 각국에서는 나름대로 자국어권 중심으로 한 절차들이 만들어져 있다. 우리나라의 경우 전기음향이나 건축음향분야에서 실내의 음성전달 성능 평가라는 차원에서 일부 제안되어 쓰이고 있고 언어치료의 관점에서도 준비되고 있으나 아직 표준적인 방법으로 널리 쓰이고 있는 것은 없으므로 이에 대한 종합적인 검토와 함께 두루 쓰일 수 있는 표준적인 절차 마련이 시급하다.

음절명료도는 정보전달의 가장 기본적인 성능을 표시하는 평가척도로서 각각의 음운의 이해성을 진단적으로 평가하는 것이다. 음절을 이용할 경우는 단어에 비해서 비교적 평가시험의 규모가 작아도 되므로 합성시스템의 문제점을 진단적으로 평가할 때 효과적인 방법이다. 시험을 위해 준비할 음성은 목적 및 평가규모에 따라 단음절, 2음절 또는 그 이상을 대상으로 하여 다양한 환경이 포함되도록 고려된 목록의 설계가 필요하다.

2) 단어이해도

가 고립단어에 의한 평가

평가에 이용할 단어를 선택할 때, 단어의 길이, 중요도, 포함되는 음소의 종류 등을 고려하여야 한다.

평가용 단어세트를 구성할 때 고려할 사항은

- 우리말 단어 모집단 전체의 성질을 반영할 것.
- 리스트는 고정하지 않고 평가 때 마다 새로운 단어세트를 추출할 수 있을 것.

이를 위하여 기계적으로 단어세트를 선택하는 자동 프로그램의 필요하다.

나. 문법에 맞는 무의미 문장에 의한 평가

문장중의 단어 음성이해도로서 문장의 의미의 영향을 받지 않도록 문법적으로는 타당하나 문장으로서의 의미가 부자연스런 문장을 사용하는 방법도 있을 수 있다

3) 문장이해도

문장 이해도는 문장으로 된 합성음을 듣고 그 문장을 올바르게 들었는가를 평가하는 것이지만 문장에 따라서는 올바르게 들는데 장애가 되는 요인이 다양하다. 따라서

- 평가에 사용되는 문장은 어떤 것이 바람직한가
- 질문과 응답방식은 어떤 것이 좋은가
- 평가문장과 질문의 구체적인 예

등이 세부적으로 검토되어야 한다.

3. 자연성의 평가

자연성의 개념자체가 여러 품질 요인이 복합되어 있

다고 생각되므로 자연성 평가에 대해서는 종합적으로 다룰 필요가 있다. 다음과 같은 항목의 검토가 필요하다고 본다.

- 대상이 되는 문장의 종류, 문장 길이 및 문장 선정

방법

- 제시순서
- 합성을 청취방식 (스피커, 헤드폰, 전화 핸드셋 등)

- 평가항목

- 회답방식

그러나 자연성의 평가는 인간의 심리적인 평가를 기반으로 이를 객관화 시키는 과정이므로 체계적인 기초연구가 뒷받침되어야 할 것으로 보인다. 따라서 이를 객관화하여 품질평가의 기준으로 하기에는 다소 시간이 걸릴 것으로 보인다.

4. 종합평가

음성합성시스템은 사람과 기계간에 인터페이스라는 시스템 요소로 위치하므로 합성시스템 자체의 기술적인 달성도를 평가하는 진단적인 평가(이해성, 자연성 등) 이외에 이용목적, 이용환경, 이용방법, 이용자 요인 등을 고려한 종합적인 평가가 필요하다. 그 평가 방법을 아직 정립되어 있다 할 수 없으나 될 수 있는 한 실제사용에 가까운 상태로 주관적인 인상을 종합적으로 평가하여 이용목적에 대한 적합성을 평가할 필요가 있다. 현재로서 가능한 방법은 스케일 평가척도를 사용하여 필요한 항목들을 개별 평가하여 종합하는 방법이 가능할 것이다.

1) 평가대상

평가 대상으로 생각할 수 있는 항목은 다음과 같다

- 이해성 : 합성음을 듣고 그 이해성 및 읽기에러 유무를 평가한다.
- 음질 : 음질 및 목소리의 질을 평가한다.
- 시간적인 요인 : 우리말다운 리듬감, 발성속도 연결의 자연스러움을 평가한다.
- 억양 : 자연성과 유창성을 평가한다.
- 종합적인 양호도 : 인간의 목소리다움, 합성음에 대한 선호도, 합성음 전체에 대한 좋고 나쁨을 평가한다.

- 이용 목적에 따른 적합성 : 합성된 목소리의 질, 발성속도 및 억양이 이용목적에 적합한지의 여부, 피로감을 느끼지 않고 기분 좋게 느껴지는 지, 합성음성이 이용목적에 적합한지여부, 저해요인은 없는지 등을 평가.

2) 평가방법

응용에서 실제 사용하는 텍스트나 또는 이와 동일한 문체의 텍스트를 평가용 문장으로 한다. 이 평가용 음성을 듣고 각 평가항목에 따라 스케일평가 척도로 평가한다. 이때 평가척도의 단계는 외국의 사례들로 7단계로 하는 것이 일반적으로 보인다.

예를 들면,

- 3. 아주 나쁘다.
- 2. 꽤 나쁘다.
- 1. 조금 나쁘다.
- 0. 보통이다.
- 1. 조금 괜찮다.
- 2. 꽤 좋다.
- 3. 아주 좋다.

III. 한국어 TTS 시스템 평가를 위한 과제

지금까지의 검토를 통해서 한국어 TTS 시스템의 객관적인 평가방법을 마련하기 위하여 우선적으로 추진할 사항들을 정리하여 보면 다음과 같다.

1. 텍스트 해석의 평가

- 읽기에러 평가를 위한 평가단위의 논의가 필요.
- 장단, 강세, 포우즈 등의 에러 판단 기준 및 방법.
- 진단적인 평가를 위해서는 형태소, 구문, 의미해석의 평가방법의 검토도 필요.
- 이러한 평가를 측면 지원하기 위하여 음운 및 운율 정보가 잘 표기된 텍스트 및 음성 코퍼스의 확보.

2. 이해성 평가

- 평가용 시료, 시험방법, 평가자, 시험조건 등의 기록 등을 포함하여 기존의 국내 제안 사례들을 참고하여 응용 분야별로 두루 쓰일 수 있는 공통적이고 통일된 방법의 마련.
- 시험용 음절, 단어 및 문장 자동 선정을 위한 알고리즘 및 선정 자동화 프로그램의 마련과 이를 위한 텍스트 코퍼스의 확보.

3. 자연성 평가

이에 관해서는 좀 더 기초적인 연구가 축적될 필요가 있다고 본다.

4. 종합평가

- 종합평가에 적절한 공통적인 평가항목의 설계.
- 이에 따른 질의 응답방식 등 평가방법의 결정.
- 질문종합평가 척도의 제시.
- 아울러 응용분야를 유형화하여 실제 사용 환경에서 종합적으로 평가할 수 있는 시범적인 절차의

제시.

IV. 결론

지금까지 TTS 방식의 우리말 음성합성시스템에 대한 객관적인 성능평가법 마련을 위하여 고려할 사항들을 살펴보았다. 최근 음성정보기술산업지원센터(SiTEC)의 요청 및 지원에 의해 대한음성학회 내에 전문가 그룹이 구성되어 음성인식도 포함하여 합성시스템의 평가법 마련을 위한 활동을 시작하려 하고 있다. 관심 있는 전문가들의 참여가 있기를 바라고 특히 텍스트 해석의 경우는 자연어처리 전문가, 이해성 평가는 국어학, 음성학, 심리학 등의 전문가와 음성공학 연구자들의 공동 노력이 필요하다고 생각한다. 특히 이해성의 평가법은 음성공학만의 요구가 아니고 어학 교육, 전기음향, 건축음향, 심리음향, 음성통신, 언어치료 등 응용분야가 다양하므로 자체에 학제적인 공동 협력을 통하여 공동으로 사용할 수 있는 통일된 절차가 마련될 수 있기를 기대한다.

참고문헌

- [1] Dafydd Gibbon et al, "handbook of standards and resources for spoken language systems", Mouton de Gruyter, 1997.
- [2] Makoto Akabane et. al, "Standars of speech synthesis system performance evaluation methods", J. of Acoustical society of Japan Vol 59-12, 2003(in Japanese).
- [3] S. Itahashi, "Guidelnes for Japanese speech synthesizer evaluation", Proc. LREC2000 Vol 2, 2000.
- [4] 이용주, "음성합성 및 인식의 성능평가와 음성 DB", 제9회 음성통신 및 신호처리 워크샵 논문집, 1992.8
- [5] 강경옥, 강성훈, "음성품질평가에 관한 CCITT 연구동향", 9회 음성통신 및 신호처리 워크샵 논문집, 1992.8
- [6] 이만영 외, "명료도 평가법 연구ETRI 위탁연구보고서", 1990.11
- [7] 최진태, "한국어 명료도에 관한 연구", 채신부 진기통신연구소 보고서, 1969.
- [8] 김순협 외, "합성음의 이해성 평가를 위한 단어표 및 실험법에 관한 연구", ETRI 위탁연구보고서, 1990.11
- [9] 조철우, 김경태, 이용주, "무의미 단어에 의한 규칙 합성음의 평가 및 진단법에 관하여", 10회 음성통신 및 신호처리 워크샵 논문집, 1993.8