# 韓國語 規則 音聲 合成 시스템의 具現

손 영 택, 김 용 갑, 松本 達郎
한국후지쯔(株), 후지쯔(일본) 연구소(株)

# The Design and Implementation of Korean Text-to-Speech Conversion System on a Rule-Based Framework

Yungtaek Son and Yongkap Kim
FUJITSU KOREA LTD.
Tatsuro Matsumoto, K. H. Loken-Kim
FUJITSU LABORATORIES LTD.

## 요 약

본고는, 한글 한자가 혼용된 입력 텍스트를 음성으로 변환 출력하는 포르만트 음성 합성 방식 즉, 한국어 규칙 음성 합성(이하에는 KTTS [Korean Text To Speech System]이라고 함)의 전반적인 처리 흐름에 대하여 소개한다. 특히, 입력 텍스트에 있어서, 한자 또는 각종 부호의 한글 변환 기능, 음성 출력용 문법 정보 추출에 필요한 입력문의 해석 및 구문 경계 설정 기능, 또한 음소 기호 변환 및 파라메터 값 생성과 변경 처리 기능을 중심으로 설명하고자 한다. 또한 본 시스템의 완성과 더불어 실시하였던 청취 실험 평가 결과에 대하여 덧붙이겠다.

## Ⅰ. 序論

현재, 기계와 인간과의 정보 전달 매체 개발에 있어서, 음성의 인식 및 합성의 연구가 활발히 진행되고 있다. 이와 때를 같이 하여, 한국후지쯔(株)는 일본 후지쯔 연구소(株) (가와사끼 소재)에서 개발중인 다언어 음성 합성 시스템 개발[1]의 一環으로 1990년 3월부터 2년에 걸쳐서 KTTS를 개발하였다. 본 KTTS는 입력된 문장을 준비된 갖가지 규칙을 이용하여서 합성 음성으로 변환하여 주는 기능을 하며, 이를 위한 규칙에는 전처리 규칙, 구문 경계 추출 등에 필요한 언어 처리 규칙, 음소 변환 규칙, 음향파라메터 생성 규칙 등이 이용된다.

짧은 기간에 본 시스템을 프로토타입으로 完成할 수 있었던 理由는, 한국어는 발음학상 英語와 類似하고, 언어 및 문법적으로는 일본어와 유사한 면이 있기 때문에, 필요한 경우에 기존의 관련 기능들을 응용하고, 한국어의 언어학적, 음성학적 독특한 특징을 추출하는 기능의 음성 합성용 프로그램을 작성하여, 각 機能간 전체 인터페이스를 조정 결합함으로써 가능하였다. 개발 환경으로써는, UNIX 시스템 상의 SUN Sparc Station을 利用하였고, 시스템의 전체 크기는 220Ks이며 음성 합성에 걸리는 시간은 실시간의 5배정도이다. KTTS의 전체 처리 흐름은 그림1에 제시한다.

本考의 구성으로는, 2章에서 KTTS의 시스템 구성에 대하여, 3章에서 후지쯔 연구소(株)에서 實施한 실험 평가 結果, 問題點 및 개선점에 대하여, 4章에서는 結論에 對하여 敍述하고자 한다.

## Ⅱ. KTTS의 시스템의 構成

KTTS시스템은 언어 처리 및 음향 처리에 있어서, 규칙을 이용하여 입력문장을 음성으로 출력하는 포르만트 음성 합성용 프로그램이다. 본 시스템은 전처리부, 언어

처리부, 음소 및 음향처리부, 합성 처리부로 크게 구분한다[2].

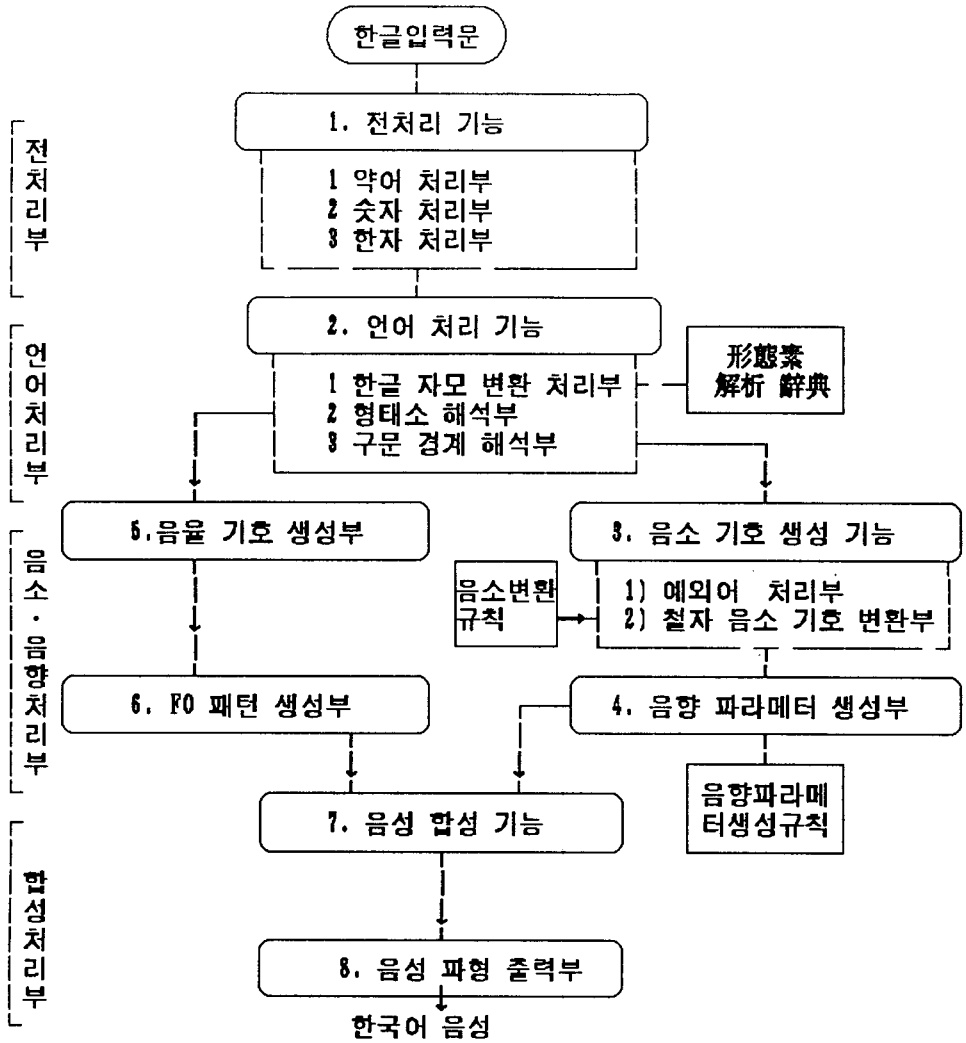


그림1. KTTS의 전체 처리 흐름

현재, KTTS는 약 50,000단어의 형태소 해석 사전, 650개의 음소 변환 생성 규칙과 600개의 음향 파라메터 생성 규칙을 포함하고 있다.

본 장에서는, 그림1의 각 처리부에 있어서, 기능, 방법 및 결과 등에 대하여 설명한다. 예를들면, 숫자 및 알파베트가 혼재된 한글 입력문을 대상으로 갖가지 처리 기능을 수행한결과로써, 合成音이 출력되는 處理과정을 上記 各 처리단계를 이용하여 說明하겠다.

## 1. 전처리 기능

입력문 중 혼재되어 있는 숫자, 알파베트 및 한자를 대상으로 숫자와 알파베트는 발음상의 한글식 표기로, 한자는 한글 표기로 변환시키는 기능을 하며, 언어 및 음향 처리를 위한 준비 단계로써 약어, 숫자, 한자 처리부로 구성되어 있다.

1) 略語 처리부
   입력 문장內의 약어, 기호 등을 對象으로 약어 및 기호 변환 규칙을 이용하여 발음상의 한글식 표기로 변환 출력하는 기능을 한다.
2) 숫자 처리부
   입력 문장內의 숫자를 포함하고 있는 단어(숫자, 연호)를 대상으로 숫자의 한글 변환 규칙을 이용하여 발음上의 한글식 표기로 변환 하는 기능을 한다.

3) 한자처리부

입력 문장내의 한자를 대상으로 한자의 한글 변환 테이블을 이용하여 한글로 변환하는 기능을 한다.

## 2. 言語 處理 機能

전처리 기능에서 변환된 한글 문장을 입력받아서 음향 처리용 음율 제어 정보를 추출하는데 필요한 정보를 얻기 위하여, 입력문 內 단어의 품사 정보, 문장의 문법구조 정보를 얻어 내는 기능을 한다. 한글은 대개 초성, 중성 및 종성으로 이루어지며, 경우에 따라서는 초성, 중성으로도 표시 가능하다. 이러한 한글에는 형태소間 경계가 존재하며, 動詞의 경우에 한글의 음절內 형태소 경계가 있는 경우가 있기 때문에 입력된 한글을 각각의 자모로 분리하고 있다. 本 처리에서는 한글의 자모 변환, 형태소 해석 및 구문 경계 해석 處理에 대하여 敍述한다.

1) 한글 자모 변환 처리부

한글은, 동사의 경우 한글의 음절內 형태소間 경계가 존재하므로, 입력 한글을 자모로 분해할 필요성은 이미 언급하였다. 이러한 필요성에 의해, 本 처리부에서 한글 입력문을 자모로 변환하는 처리 기능을 행한다.

2) 형태소 해석부

형태소 해석에 사용된 방식으로는 최장 일치법을 利用하고 있다. 자모로 분리된 입력문을 형태소 해석하기 위하여 형태소 해석 사전, 단어間 인접 접속표, 단어의 인덱스 정보를 참조한 後, 各 단어의 품사 정보를 얻는다. 형태소 해석 사전에는 約 50,000個 이상의 단어가 등록되어 있고, 인접 접속표는 128個의 품사를 정의할 수 있는 128 x 128 메트릭스로 이루어졌다. 478個의 문장을 이용한 형태소 해석 처리 결과는 95.6%의 解析率을 나타내었다. 특히, 동일 표기 단어의 해석 時 오류 발생率이 높았다. 解析 관련 처리의 一例는 그림2와 같다.

入力文: 나는 밥을 먹는다.

字母變換 | ㄴ ㅏ | ㄴ _ ㄴ | ㅂ ㅏ ㅂ | ㅇ _ ㄹ | ㅁ ㅓ ㄱ | ㄴ _ ㄷ ㅏ | . |

品詞: 代名詞 助詞 名詞 助詞 動詞語幹 敍述形語尾 .

그림2: 형태소 해석 관련 處理의 一例

3) 구문 경계 해석부

문장內 構文과 構文間의 정의는 抑揚의 형태를 생성하는데 중요한 역활을 한다. 보다 자연스러운 억양을 제어하기 위하여 형태소 해석의 결과를 이용하여 문장내의 구문 경계 정보를 추출하며, 또한 음향 처리 時 중요하게 영향을 미치는 호흡 관련 정보의 추출 기능도 本 처리에서 행한다.

구문 경계 해석의 기본 사상은 하나의 한글 구문은 1 個의 자립어와 1個 이상의 부속어로 구성된다[3]. 그림3에서는 입력 문장중 2 個(A, B)의 구문을 나타내며, X는 A구문의 마지막 형태소 즉 부속어를, Y는 B구문의 첫번째 형태소인 자립어를, Z는 B구문의 마지막 형태소 즉 부속어를 나타내고 있다. 本 처리에서는 8개 자립어, 55 개의 부속어를 정의하고 있으며, 표1에서 보는 바와 같이 구문간 의존 관계를 7개로, 또한 구문간 의존 강도를 3개 level로 정하고 있다.

주어진 구문간의 관계를 정의하기 위해서는 구문간 의존 관계 및 의존 강도가 정의된 XYZ메트릭스 테이블을 참조하고 있으며, 구문 경계 해석의 기본 사상은 그림3에 나타냈다. XYZ메트릭스 테이블 작성은 구문간의 관계를 파악한 후 手작업으로 정의하였다.

구문 의존 관계 점수의 계산 방정식은 (1)과 같다. Sij는 구문 의존 관계 점수이고, Rij는 i번째 구문과 j번째 구문間 XYZ메트릭스 테이블上의 의존 강도를 나타낸다. A와 B는 임의의 常數이다.

$$Sij = A * Rij + B * (j - i) \qquad (1)$$

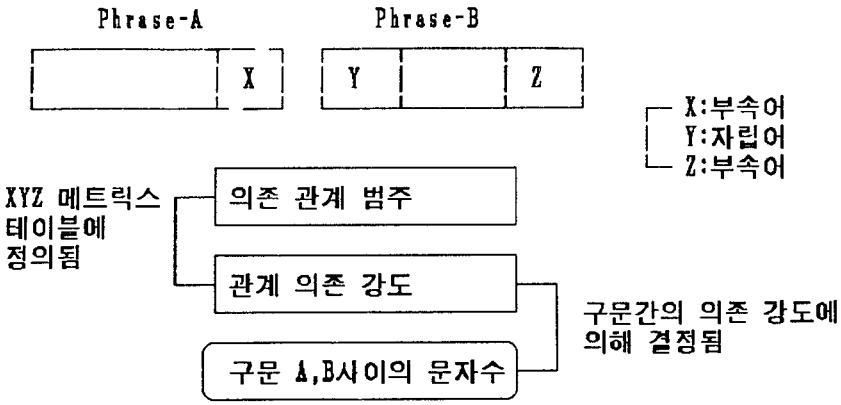


그림3: 구문 경계 해석의 기본 사상

표1:구문간 의존 관계 범주

| NO | 관계 범주 |
|---|---|
| 0 | 관계가 없다 |
| 1 | 주어-술어 구문 |
| 2 | 부사-술어 구문 |
| 3 | 형용사-명사 구문 |
| 4 | 병렬구문 |
| 5 | 보조-수식 구문 |
| 6 | 독립어 구문 |

구문 의존 관계 점수가 위의 방정식(1)에 의해 계산이 되면, 구문 경계 해석 기능은 각 구문간의 의존 관계 점수가 가장 높은 관계를 선택하여 구문 관계로 결정한다. 그림4에서 보는 것처럼, 여러 가지의 의존 관계상에서 크로스되지 않는 범위내에서 S14, S24, S34, S45가 최고 의존 점수로써 선택된 구문간의 경계이다. 本 구문 경계 해석에 대한 평가는 형태소 해석에서 사용했던 478個의 평가문을 대상으로 평가를 실시하여 91.8%의 정확성을 얻었다.

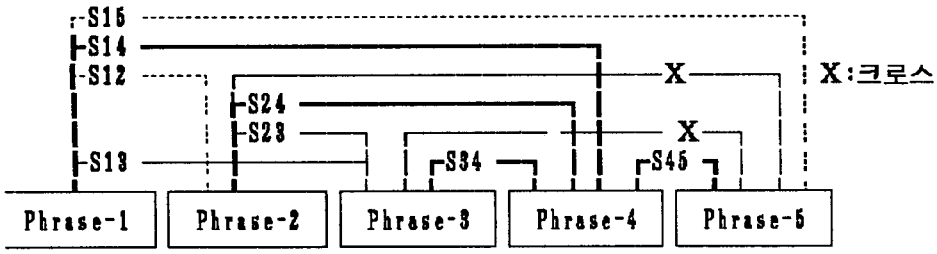


그림4:구문간의 구문 경계 설정 一例

### 3. 音素 記號 生成 機能

음향 처리를 하기 위한 준비 단계로써, 언어 처리된 한글 자모 입력문을 음소 기호列(음성 관련 국제 표준 기호)로 변환하는 기능을 한다. 음소 변환 규칙은 1988년 교육부에서 제정한 표준 한국어 발음에 기초하여 작성하였다[4]. 本 기능에는 예외어 처리 및 철자 음소 기호 변환 처리를 행하고 있다[5].

1) 예외어 처리부
일반적인 발음 규칙을 적용 받지 않는 합성어 및 파생어를 대상으로 해당 음소 기호를 예외어 사전에 등록하여 놓고, 등록된 사전을 탐색하여 음소 기호로 변환하는 기능을 한다.

2) 철자 음소 기호 변환부

예외어 처리를 적용 받지 않은 단어를 대상으로, 그림5에서 보는 바와 같이 각각의 문자열을 중간표현인 알파베트로 바꾸고, 음소 변환 규칙을 이용하여 음소 기호로 변환을 행한다. 음소 변환 규칙은 각종 음운 현상(자음접변, 구개음화, 말음법칙등) 및 발음 규칙이 적용되어 있다. 본 규칙은 음소 생성의 규칙 형태로 표현되고 있으며, 그림5에서의 밑줄 부분(_)은 변환될 철자의 위치를 나타낸다. 本 處理部에서는 39개(15개 파열음계 4개의 파찰음계, 4개의 마찰음계, 3개의 비음계, 2개의 유음계, 9개의 모음과 2개의 반모음)의 음소 기호로 분류하였다.

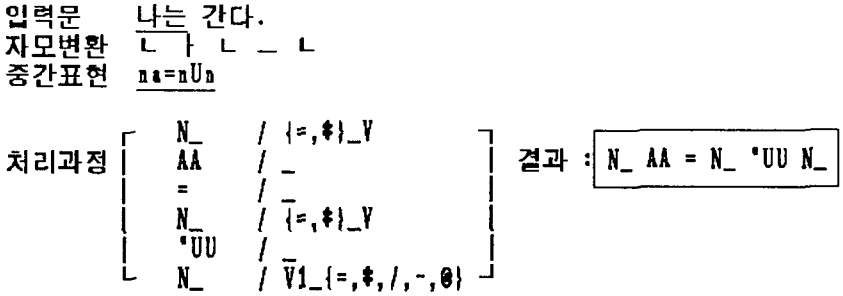


그림5:음소 변환 기능 결과의 一例

## 4. 음향 파라메터 생성부

음소 기호열로부터 연속된 시간의 음향 파라메타를 생성하는 기능을 한다. 포르만트 주파수, 음원 진폭, 음원 증폭 등의 20개의 시계열 음향 파라메터로 구성되어 있으며, 이들은 KLATT형 포르만트 합성기에서 취급된다. 이 파라메터의 값은 1732개의 한글 철자로부터 파형 분석으로부터 추출되었고, 이 값은 음소 길이, 음원 증폭, 모음과 자음의 음향 파라메터 관련으로 음향 파라메터 생성 규칙 작성에 참조된다[6]. 현재는, 600개의 파라메터의 설정 및 수정 규칙 등이 이용되고 있다. 이 규칙들은 해당 음소열에 음향 파라메터의 값을 할당하고 수정한다. 음향 파라메터 생성의 처리 과정을 아래에 소개한다.

【음향 파라메터 생성 과정】

1) 데이터의 초기화 : 각종 음향 파라메터의 데이터를 초기화한다.

2) 규칙을 읽어 들임 : 음향 파라메터 생성 규칙을 읽어 들인다.

[음향 파라메터 생성 규칙의 종류]

(1) 시간 길이 설정 규칙
: 각 음소의 고유 시간 길이를 설정한다.

(2) 시간 길이 수정 규칙
: 음소 환경에 의해 변화하는 기본 음소의 시간 길이를 수정한다.

(3) 엑센트 및 문 전체의 강세 설정 규칙
: 품사 정보를 이용하여 강세의 크기를 설정하고, 구문 경계 정보(호흡 기호 등)에 기초하여 문 전체의 강세를 결정한다.

(4) 음원 진폭 설정 규칙
: 음원의 고유 진폭을 설정한다.

(5) 모음 파라메터 설정 규칙
: 모음에 관계된 음향 파라메터를 설정한다.

(6) 자음 파라메터 설정 규칙
: 자음에 관계된 음향 파라메터를 설정한다.

(7) 음원 진폭 수정 규칙
: 음소 환경에 의해 변화되는 음원 진폭을 수정한다.

(8) 파라메터 수정 규칙
: 음소 환경에 의해 변화되는 음향 파라메터를 수정한다.

3) 변별 속성 테이블을 읽어 들임 : 음소 마다에 관련된 변별 속성 테이블을 읽어 들인다.

4) 단어 및 음소 frame 생성 : 입력 단어 및 음소 마다의 데이터 격납용의 frame을 생성한다.

5) 규칙의 적용 : 각 규칙을 음소 frame에 적용하고, 각 음소 frame내의 적용결과에 따라서 음향 파라메터의 시계열을 생성 수정한다.

6)음소 frame의 연결: 각 음소 frame을 연결하고, 문 전체에 따르는 음향 파라메터 시계열을 작성한다.

7) 음향 파라메터 시계열 출력 : 음향 파라메터의 시계열을 출력한다. 생성된 음향 파라메터의 예를 표2에 나타낸다.

표2: 음소의 시계열 음향 파라메터 값의 一例

| TIME (msec) | AV (dB) | AH (dB) | F1 (Hz) | F2 (Hz) | F3 (Hz) |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 780 | 1100 | 2800 |
| 5 | - | 40 | - | - | - |
| 25 | 48 | - | - | - | - |
| 50 | 53 | 0 | - | - | - |
| 70 | - | - | - | 1100 | - |
| 95 | 48 | - | 780 | 1326 | 2800 |
| 120 | - | 0 | - | - | - |
| 155 | 50 | - | 375 | 2052 | 2890 |
| 190 | 55 | 30 | - | 2234 | - |
| 230 | - | - | - | - | - |
| 255 | 50 | 30 | - | - | - |
| 290 | 0 | 0 | 375 | 2234 | 2890 |

## 5. 음율 기호 생성부

언어 처리 기능으로부터 얻어진 품사 정보 및 구문 경계 정보를 이용하여 문장 전체의 억양 및 운율 관련 제어 기호를 생성한다.

음율 기호 생성 방법으로는 첫째로, 구두점에 의한 억양의 개시점 및 호흡에 관계한 부분에 음율 기호를 할당한다. 둘째로, 2個의 연속한 구문사이에서 문법적인 연관 관계를 고려하여 구문과 구문사이에 음율기호를 둔다. 마지막으로, 2個의 음율기호 사이의 음절 갯수가 최대 음절수 즉, 8이상이 될 경우에 음율기호를 부가한다.

## 6. F0패턴 생성부

F0패턴 생성부에서는 후지사끼 先生의 F0패턴 모델을 이용하여 음율기호로부터 기본 주파수(F0) 패턴을 만든다[7]. 그림6에서 기본 주파수 패턴의 一例를 보여 주고 있다. 한국어를 위한 후지사끼 모델의 적용 결과는 표3에 제시하였다. F0용 파라메터 값을 얻기 위하여 韓國人 남자 아나운서의 音聲을 分析 合成하는 TOOL을 이용하였다.

표3: 후지사끼 모델을 적용한 韓國語用 F0 파라메터 값

| 파라메타 | a | b | c | P0 | P1 | P2 | P3 | A1 | A2 | F0min |
|---|---|---|---|---|---|---|---|---|---|---|
| 값 | 2.05 | 27.94 | 0.9 | 0.42 | 0.60 | 0.23 | 0.10 | 0.30 | 0.25 | 77.15 |
| 始作시간(msec) | | | | 150 | -340 | -150 | -100 | -60 | -60 | |

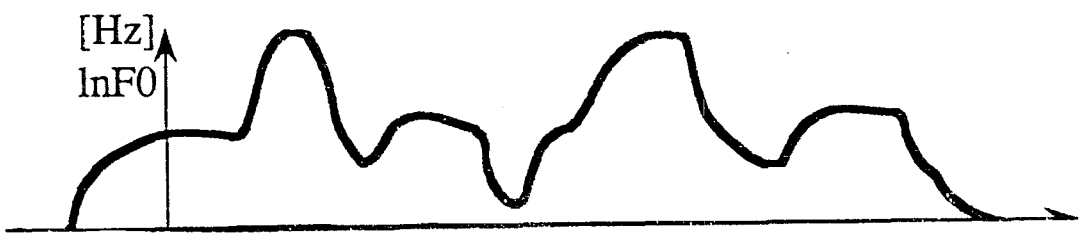


그림6:F0 형태의 一例

7. 음성 파형 출력부

합성음 및 음성 파형을 생성하는 모듈로써, 수정된 KLATT형의 합성음 필터를 사용하였다. 이 필터는 "A SOURCE WAVEFORM MODEL"을 사용하는 ROSENBERG SOURCE WAVEFORM을 이용한다. 각 파라메터는 가상 스펙트럼(1/F)의 무작위 숫자를 사용하고 있다. 음성 파형 출력 기능은, 기존의 영어 음성 합성 시스템의 파형 출력부를 修正없이 사용하고 있다.

이상의 처리 과정을 거친 후, 한글 입력문이 KTTS에 의하여 음성 출력이 가능하게 된다.

# Ⅲ. 실험평가

본 실험에는 1名의 여성과 2名의 남성 한국인이 참여하였고, 합성음에 경험이 없는 피험자들로 구성되었다. 제시된 실험 句節은 국민학교 1학년에서 3학년의 국어 교과서에서 발췌하였고 총5個이며, 각각은 7에서15개의 문장으로 이루어졌다.

각 문장은 KTTS에 의하여 합성되었고 오디오 테이프에 녹음하여서, 각 句節로부터 발췌한 질문지(총25문)와 함께 피험자에게 제시되었다. 질문지는 4지선택의 형식으로 테이프상의 정보와 동일한 순서로 질문이 준비되었다.

피험자들은 本 실험 실시前에 KTTS에 의하여 합성된 음을 얼마나 이해하고 있는가 等의 실험 평가 목적을 설명듣는다. 그리고, 청취하였던 문절의 내용에 대한 질문에 대하여 4가지 중 택일하게 된다. 녹음 테이프는 2번 제시되며, 합성음을 들은 後에 질문 내용을 펴 볼 수 있었다. 대신에, 합성음을 듣는 동안에는 간단한 메모를 허락하였다. 평가실험의 절차와 결과, 문제점 및 改善点을 다음에 열거한다.

1. 실험 평가의 절차와 결과

1) 평가의 절차
- 평가용 자료의 수집
  (국민학교 1,3학년용의 국어 교과서로부터 무작위로 선별한 내용(5 Parts)을 중심으로 각 Part는 10여개의 예문으로 구성되었다.)
- 평가용 피험자의 선정
  (한국어를 이해할 수 있는 Native Speaker인 3名의 한국인을 피험자로 선정하였다.)
- 평가용 자료의 합성음 생성
  (수집한 평가용 자료를 대상으로 합성음을 작성하였다.)
- 평가 실험의 실시
  (각 Part별 합성음을 2番 청취한 후 이해도를 측정하는 평가 실험을 실시하였다.)

2) 평가의 결과

평가 실험의 결과는 표4에 나타냈다.

표4: 실험 평가 결과표 (단위:文)

| 문 절 | SK | LK | JS | 이해도(%) |
|---|---|---|---|---|
| 1 | 3 | 4 | 4 | 73.3 |
| 2 | 5 | 4 | 3 | 80.0 |
| 3 | 3 | 4 | 3 | 66.7 |
| 4 | 5 | 5 | 5 | 100.0 |
| 5 | 5 | 5 | 5 | 100.0 |
| 평 균 | 84. | 88. | 80. | 84.0 |

### 2. 평가上의 문제점

本 실험에서 사용한 질문지 및 결과에 대하여 객관성을 論하기에는 2가지면에서 문제점이 있다.

1) 피험자가 충분하지 못한 점.
   - 단지 3명에 의한 실험 평가의 결과에 대한 신뢰성은 낮다.
2) 이해도의 평가만을 실시한점.
   - 주로, 이해도 평가를 측정하였기 때문에 추출한 합성음 및 질문지의 내용이 합성음의 명료도 등의 측정에 객관성이 부족하였다.

따라서, 보다 많은 피험자를 대상으로 보다 객관적으로 합성음을 평가할 수 있는 평가 실험을 실시할 필요가 있다.

### 3. 개선점

실험 결과 및 피험자로부터 얻은 KTTS의 개선점에 대하여 다음에 열거한다.

1) 합성음의 음질의 개선의 필요성
   - 합성된 음성을 어느 정도 이해하였는가에 있어서, 합성음의 음질이 가장 중요한 포인트가 된다. 本 KTTS의 음질은 아직 개선의 여지가 많다. 특히, 자음에 있어서, 파열음간의 구별이 어려우며, 음질을 향상시킬 필요성을 느낀다.
2) 속도 향상의 필요성
   - 입력문의 합성음을 얼마나 빠르게 출력하느냐는 음성을 이용한 정보 전달의 응용 및 장점을 극대화 시키게 된다. 그러나, 현재, 처리 속도는 실시간의 5배 정도로 개선의 필요성이 절실하다.

## Ⅳ. 結論

3章에서 한국어 음성 합성 시스템의 평가 실험 결과 (이해도:84%) 및 문제점을 열거하였다. 本 시스템의 결과를 정리하면, 형태소 해석의 결과는 96%이고, 구문 경계 해석은 92%이다. 3名의 한국인을 대상으로 실시했던 간단한 청취 실험의 이해도는 84%의 결과로 충분히 흡족한 결과가 되지 못한다. 앞에서 언급한 문제점을 해결하고 합성음의 음질을 향상시키기 위하여 음향 파라메터 등 필요한 규칙을 수정 보완할 계획에 있다.

## 참고문헌


[1] Matsumoto, T., "A Study of the Multilanguage Text-to-System by Rules - Text-to-Phoneme Conversion -," Proc. IPSJ.,September,1986(in Japanese).
[2] Matsumoto, T.and Son, Y. T.,"Syntactic Analysis for Korean Text-to-Speech System," Proc. IEICE., March, 1992(in Japanese).
[3] Kaseda, M. et al, "Studies on pitch contour and duration control in Speech Synthesis by Rule," Proc. Acoust. Soc. Jp.,March, 1987(in Japanese).
[4] Kim, Y. K., and Matsumoto, T., "Converting Korean Text to Phoneme," Proc. IEICE., March, 1991(in Japanese).
[5] Lee, E. J.,"The Rules of Korean Spelling and Standard Language Explanation - Korean Pronunciation rules -," Daejegak publishing co., 1988(in korean).
[6] Kim, Y. K., "A Rule-Based Text-to-Speech System for Korean," Korean-Japan Joint Workshop on Advanced Technology of Speech Recognition and Synthesis, Korea,1991.
[7] Fujisaki, H., Hirose, K. and Kawai, H., "A System for Synthesis of Connected Speech:Special Emphasis on the Synthesis of Prosodic Features," Trans. of Communitee on Speech Research, Acoust. Soc. Jpn, SP85-43, 1985(in Japanese).