스무고개 게임을 위한 음성인식

노용완, *윤재선, 홍광석
성균관대학교 정보통신공학부 휴먼컴퓨터연구실, *(주)보이스텍
Tel. 031-290-7196 , HP. 019-296-0594

Speech Recognition for twenty questions game

Yong-Wan Roh, *Jeh-Seon Youn, Kwang-Seok Hong
HCI Lab, School of Information and Communication Engineering, Sungkyunkwan University
elec1004@hotmail.com, sunhci@hotmailcom, kshong@yurim.skku.ac.kr

Abstract

In this paper, we present a sentence speech recognizer for twenty questions game. The proposed approaches for speaker-independent sentence speech recognition can be divided into two steps. One is extraction of the number of syllables in eojeol for candidate reduction, and the other is knowledge based language model for sentence recognition.

For twenty questions game, we implemented speech recognizer using 956 sentences and 1095 eojeols. The results obtained in our experiments were 87% sentence recognition rate and 90.15% eojeol recognition rate.

Ⅰ. 서론

음성인식 기술의 비약적인 발전에 따라 음성인식 기술을 이용한 많은 제품들이 등장하고 있다. 그러나 대부분 적은 어휘를 대상으로 하는 고립단어 인식을 이용하는 수준이고, 핵심어 또는 문장인식의 적용 상품도 일부 소개되고 있지만 자연스러운 사용자의 발성을 인식하는 수준에는 아직 미치지 못하고 있다. 또한 인식된 문장을 이해하여 이해된 결과에 의해 판단 및 반응을 하는 것은 초보적 단계에 있으나 많은 기업과 연구소에서 이를 위한 연구가 활발히 진행되고 있다.

본 논문에서는 자체 제작한 음성인식기와 합성기를 이용하여 스무고개 게임을 구현하였다. 일반적으로 스무고개 게임은 영역의 제한이 없지만 본 논문에서는 제한된 영역으로 구현하였다. 인식의 기본 단위는 어절로 하였고, 인식 후보를 줄이기 위해서 음절의 개수를 추출하여 음절수에 따라 후보를 제한하는 알고리즘을 적용하였다. 문장인식 시 언어모델을 구성하여 후보의 수를 줄여서 인식하도록 하였다.

스무고개 게임에는 임의의 동물 또는 곤충을 선택한 후, 956문장 유형 중 하나를 화자가 발성하면, 발성된 문장을 인식과정을 거쳐 텍스트 창에 출력한다. 컴퓨터는 선택한 동물과 관련된 답을 출력하며 사용자가 동물의 이름을 맞출 수 있도록 구성하였다. 음성 인식기는 VCCV (Vowel + Consonant + Consonant + Vowel) 기반의 화자독립 가변어휘 음성인식기[2,5]를 사용하여 구현하였다.

본 논문에서 구현된 스무고개 게임에는 음성인식과 answer table을 이용하여 정답을 제공 할 수 있도록 한다. 스무고개 게임은 956문장으로 구성 되어있다. 41개의 동물, 곤충으로부터 구성되어 있지만 다른 분야의 단어들도 적용 가능하도록 설계되었다.

Ⅱ. 스무고개 시스템

어휘독립 인식기를 말놀이 중 가장 지능적으로 발전된 형태의 놀이인 스무고개 게임에 적용하였다. 956 문장을 의미가 유사한 것들끼리 묶어 483개의 군을 구성하여 각 동물 및 곤충에 대해 답을 구하여 저장하였다. 스무고개 시스템의 개념도를 그림 2.1에 나타내었다.

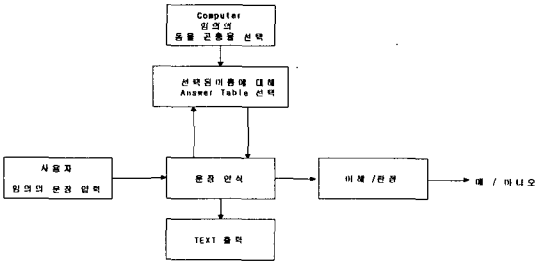

그림2.1 스무고개 시스템의 개념도

스무고개 시스템이 시작되면 컴퓨터는 임의의 이름을 선택하고, 선택된 동물 또는 곤충의 answer table을 가

져온다. 사용자는 956개의 문장 유형 가운데 한 문장을 선택하여 발성한다. 문장 인식 부에서 문장을 인식한 후, 인식된 결과의 어절 열을 text창에 나타내고, 인식된 문장이 속하는 군으로 가서 질문에 대한 답을 받아와서 결과를 합성음으로 /예/ 또는 /아니오/라고 들려주도록 구성하였다. 그림 2.2는 입력된 문장을 인식하는 과정을 나타내고 있다.

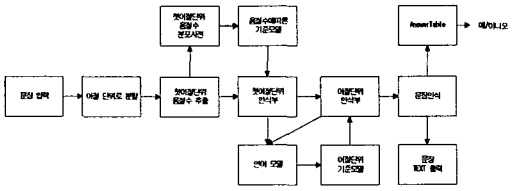

그림 2.2 스무고개 시스템 구성도

문장의 발성은 어절사이에 약간의 무음이 있는 듯한 낭독체로 발성하였으며, 문장 인식 부는 먼저 들어온 문장을 어절 단위로 분할한다. 어절과 어절의 무음 구간을 이용하여 어절을 구별하였으며, 첫 어절은 음절 개수 추출 처리를 하여 음절의 개수를 구하도록 하였다. 첫 어절의 후보는 결정된 음절 개수의 후보만을 비교하여 인식을 하였으며, 인식된 어절은 언어모델 규칙에 의해 다음의 어절 후보를 결정하고, 두 번째로 분할된 어절과 인식하도록 시스템을 구성하였으며, 문장이 완성되면 answer table을 참조하여 그 결과를 들려주고, 인식된 문장을 텍스트로 출력하도록 하였다.

Ⅲ. 스무고개를 위한 음성인식

스무고개의 질문은 50명에게 실제로 스무고개를 하도록 하여 받아 적은 후 정리해서 1,839개의 문장을 얻었으며 동일한 질문이나 상식 밖의 질문을 추려내 956개의 문장을 선별하였다.

스무고개 시스템은 41개의 동물, 곤충의 이름이 등록되어져 있으며 956개의 문장을 이용하여 사용자가 선택한 동물의 이름을 맞추도록 구성하였다. 놀이방법은 컴퓨터가 임의의 동물, 곤충을 선택한 후 사용자로부터 956개의 문장 가운데 원하는 한가지 질문을 받아 드려 문장을 인식하고, 선택된 동물과 관련이 있으면 /예/, /아니오/라고 합성음을 들려준다. 사용자는 스무 번안에 컴퓨터가 선택한 답을 맞추어야 하며, 사용자가 정답을 알면 /답은 ***입니다/ 라는 문장으로 발성하여 정답을 맞춘다. 등록된 동물과 곤충의 이름은 표 3.1에 나타내었다.

3.1 어절 단위 문장인식

본 논문에서는 텍스트가 아닌 구어음성을 위한 기본 어절 후보 규칙을 적용하여 모델을 구성하였으며, 그 규칙은 다음과 같다.

표 3.1 스무고개 정답 목록

종류	이름	개수
동물	개, 고양이, 돼지, 소, 토끼, 양, 당나귀, 말, 코끼리, 사자, 호랑이, 표범, 캥거루, 코알라, 코뿔소, 하마, 고래, 상어, 오징어, 문어, 수달, 까치, 제비, 비둘기, 참새, 닭, 앵무새, 칠면조, 오리, 원숭이, 쥐, 곰, 사슴, 뱀, 너구리, 족제비, 펭귄, 개구리	38
곤충	거미, 개미, 꿀벌	3

(1) 체언과 조사가 함께 연결된 어절 단위를 기본 후보 어절 단위로 한다. 예를 들면 /목적+으로/, /육식+을/, /모양+의/과 같은 경우이다.

(2) 체언과 체언, 체언과 기본 후보어절 단위로 연결된 각각의 어절들을 그룹으로 묶어 후보 어절 단위로 한다. 예를 들면 /이+동물은/, /육식+동물/, /서울+시내에서/의 경우 같은 경우이다.

(3) 두 음절 이하의 용언 + 체언의 경우에도 그룹핑하여 후보 어절 단위로 한다. 예를 들면 /큰+동물/, /작은+동물/, /센+동물/의 경우이다.

(4) 마지막 어절에서 '명사+입니까'인 경우에는 /명사/+/입니까/로 분리한다.

이상과 같은 규칙을 적용하지 않을 경우 어절의 총 수는 1,151개였으며, 규칙을 적용한 경우의 총 어절 수는 1,095개로 나타나 후보의 수를 줄일 수 있었다.

3.2 음절수를 이용한 인식 후보 감소

음절 개수 추출은 기준모델 작성 시 후보의 개수에 큰 영향을 미치기 때문에 중요한 과정이다. 먼저 입력 데이터로부터 에너지와 영 교차율을 이용하여 유성음과 무성음영역을 검출한다. 또한 좀더 정확한 유성음 영역을 검출하기 위해 제 1 포먼트, 제 2 포먼트가 존재하는 215Hz와 2,756Hz사이의 에너지 정보를 이용하여 안정된 유성음 영역을 추출하여 음절수를 정하였다.

스무고개 시스템에 사용된 총 어절 수는 1,095개이며, 첫 어절에 나올 수 있는 목록의 수는 561개이다. 따라서 첫 어절과 561개의 어절과 비교하게 되면 인식률의 저하 및 인식 시간이 상당히 소요된다. 따라서 음절개수 추출 방법에 따라 첫 어절의 음절 분포 사전을 구축을 한 다음, 입력 된 음성 데이터의 첫 어절의 음절 개수를 추출하여 선택된 음절 개수에 해당하는 단어만을 후보로 두고 인식하도록 시스템을 구현하였다.

표 3.2은 첫 어절의 후보 561개와 비교하여 음절개수에 따라 후보를 설정하는 방법에 따른 인식 후보 감소율을 나타내고 있다. 3음절을 제외하고 다른 모든 음절에서 50%의 미만의 감축율을 가져왔으므로 실제로 인

식 시간이 그 만큼 빨라지고 인식률의 향상이 기대된다.

표 3.2 첫 어절의 음절 개수에 따른 인식후보 감축율

음절개수의 종류	1음절	2음절	3음절	4음절	5음절	6음절	7음절
감축률	9.98%	47.8%	66.1%	39.0%	16.8%	6.06%	1.6%

3.3 언어 모델

현재의 어절과 다음 어절과의 구문규칙을 적용하여 구성하였으며 그 과정은 다음과 같다.

(1) 956개의 문장에 모든 어절의 후보 1,095개에 대해 index를 부여하고 reference 목록에 저장한다. 이 방식은 스무고개 시스템을 구현하기 위해 어절 단위의 기준모델을 메모리에 로딩하기 위해서 번호를 부여한다.

(2) 각각의 어절에 따라 현재의 어절과 다음에 오는 어절과의 문맥관계에 따라 table을 작성한다. table의 규칙은 먼저 현재의 어절이 속한 reference 목록의 순번 index, 다음 어절에 나올 수 있는 후보의 수, 다음 어절 후보의 table의 순번으로 하며 table의 수는 문장을 구성할 수 있는 어절이 총 5개이기 때문에 5개의 table을 구성한다.

(3) 첫 번째 어절의 table1은 후보가 reference 목록과 같은 번호를 가지며, 두 번째 어절과의 문맥관계를 저장한다. 예를 들어 문장이 한 어절로 이루어진 문장인 /무섭습니까/는 reference 번호가 171이라고 하면 171 0 -1 -1 …의 순으로 저장된다. 첫 번째 번호는 reference 목록의 번호이고 다음의 후보가 없기 때문에 두 번째 값이 0으로 나타낸다. 다음 번호들은 -1의 초기화값을 가지게 된다. 두 번째로 /답은/이라는 어절인 경우 다음의 후보가 41개의 동물과 곤충의 이름이 후보로 올 수 있으므로 다음 표 3.3과 같이 저장된다.

표 3.3 Reference 목록의 예

442	41	363	364	365	366	367	368	369	370	371	372
373	374	375	312	346	377	378	379	380	381	382	383
384	385	386	387	388	389	390	391	392	393	394	395
396	397	398	399	400	401	402	-1	-1			

442는 reference 목록의 번호이고 41은 다음 어절의 후보의 수가 41개가 있다는 것을 의미하며, 그 다음의 번호는 table2에 동물과 곤충의 이름이 저장된 번호를 나타내고 있다.

(4) 인식된 첫 번째 후보로부터 table2의 순번을 가져오면 맨 처음 번호의 값들을 불려 들여 두 번째 어절 단위와 비교하여 인식을 수행한다.

(5) 인식된 어절들을 어절 순서에 따라 text창에 표시한다. 그림 3.1은 인식 과정을 그림으로 도시하였다.

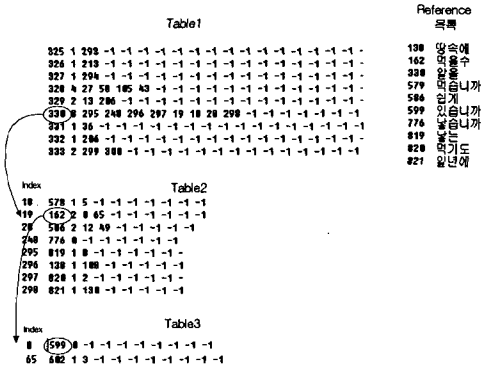

그림 3.1 인식 과정

그림 3.1는 /알을 먹을 수 있습니까/ 문장을 table에 따라 인식하는 과정을 나타내고 있다. /알을/의 reference 번호는 330, /먹을 수/는 162, /있습니까/는 599번이다. table1에서 /알을/ 인식하게 되면, table2에서 8개의 후보와 비교하여 index 19번 /먹을 수/가 인식되었고, 마지막 table3에서 index 0번과 65와 비교하여 reference 번호 599인 /있습니까/가 인식되는 과정을 설명하고 있다.

3.4 복잡도(perplexity)

복잡도는 언어모델에 의해 주어진 제약을 나타내는 것으로 일반적으로 객관적인 평가척도를 말한다. 복잡도 PP와 엔트로피 식은 다음 식(1), 식(2)와 같이 정의할 수 있다.

$$PP = 2^{-H_0} \tag{1}$$

$$H_0 = \frac{1}{N}\log_2[\mathcal{P}(w_i|w_1,\ldots,w_{i-1})] \tag{2}$$

여기서, $w_1,\ldots,w_N$는 문장의 경계정보를 가지면서 서로 독립인 평가 데이터의 어절들을 나타내고 $\mathcal{P}(w_i|w_1,\ldots,w_{i-1})$은 어절 열 $w_1,\ldots,w_{i-1}$ 다음에 발생하는 단어 w_i를 타나내는 언어모델의 확률을 나타낸다. 음절수를 적용하지 않은 경우의 복잡도는 다음과 같다.

$$H_0 = \frac{1}{5}\log_2(163068) \quad \text{(bits)} \tag{3}$$

$$PP = (163068)^{\frac{1}{5}} \fallingdotseq 11 \tag{4}$$

그러나 음절 개수 추출 알고리즘을 사용하면 복잡도가 8.47 이며 23% 감소하였다.

Ⅳ. 인식실험 및 결과

그림 4.1에 스무고개 시스템의 실행화면을 나타내었다. 먼저 목록열기를 선택하여 인식어절 모델을 메모리에 로딩을 하면, 리스트 박스에 후보 어절 목록이 나타난다. 스타트 버튼을 누르면, 컴퓨터는 임의의 문제단어를 선택하여 answer table에 등록한다. 사용자가 문장을 발성하면, 먼저 어절 분할을 한 후, 첫 어절의 음절 개수를 추출한다. 추출된 음절개수로부터 음절 개수 table에 등록되어 있는 어절후보와 비교한다.

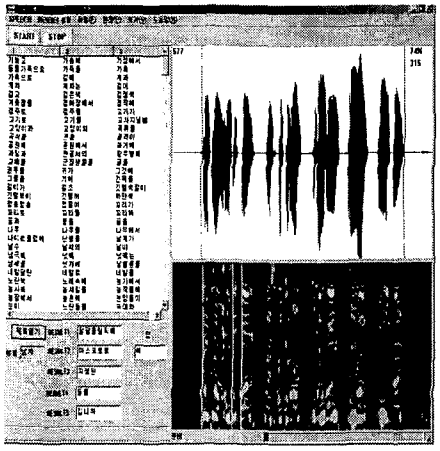

그림 4.1 스무고개 시스템의 실행화면

그림에서는 /팔팔 올림픽 때 마스코트로 지정된 동물입니까/라고 발성한 것에 대한 결과로서, 인식된 성별이 '남자'이고, 첫 어절이 5개의 음절로 분할됨을 알 수 있다. 인식된 어절의 결과가 에디터 박스에 어절에 따라 /팔팔올림픽때/, /마스코트로/, /지정된/, /동물/, /입니까/라고 쓰여지며, 프로그램이 선택된 이름의 answer table 결과를 답 에디터 박스에 나타내도록 하였다.

표 4.1 스무고개 시스템의 문장단위 인식 성능

화자	첫어절 인식률(%)	인식어절수/총어절수	총어절 인식률(%)	문장단위 인식률(%)
화자 1	87	274/299	91.64	87
화자 2	86	237/266	89.10	85
화자 3	88	228/252	90.48	88
화자 4	87	226/253	89.33	86
화자 5	92	234/250	96.30	92
화자 6	84	201/234	85.90	83
화자 7	88	223/245	91.02	88
평균	87.42	.	90.15	87.00

어휘독립 음성인식 시스템에 언어모델과 음절 개수 추출 알고리즘을 추가한 스무고개 시스템의 문장인식 성능 평가 실험을 하였다. 실험 방법은 화자 7명이 임의의 100개의 문장을 발성하였으며, 첫 어절 인식률, 총 어절 인식률, 문장 단위 인식률을 표 4.1에 나타내었다.

오 인식된 문장을 살펴보면, /털로 옷을 만들 수 있습니까/가 /털로 옷을 만들기도 합니까/로 /만들 수/가 /만들기도/와 같이 유사음절이 포함되어져 있는 문장에서 오 인식되었다.

Ⅴ. 결론

본 논문에서는 음성을 이용한 스무고개 게임의 구현에 대해 간략히 다루어보았다. 음성을 이용한 스무고개 게임은 컴퓨터가 생각하고 있는 동물이나 곤충을 사람이 맞추어 나가면서 운영하도록 하였고, 이는 이해 및 판단을 위한 응용 프로그램을 제공하는 첫 단계라고 할 수 있을 것이다.

추후에는 한정된 동물이나 곤충만 컴퓨터가 정하는 것과 화자가 이미 정해진 문장 유형을 발성하는 것이 해결하여야 할 문제이다. 이를 위하여 다른 영역 단어를 대상으로 컴퓨터가 문장을 이해 판단 할 수 있는 새로운 고안이 필요하다. 또한 컴퓨터가 질문하고 사람이 답하는 형태도 추가해야 할 부분이다.

<감사의 글>
본 연구는 한국과학재단 목적기초연구(R05-2002-001007-0)지원으로 수행되었음.

참고문헌

[1] 윤재선, 정광우, 홍광석, "무제한 단어인식 시스템을 위한 VCCV분할에 관한 연구," 한국음향학회 하계학술발표대회 논문집, 제 19권 1호, 2000.

[2] 윤재선, 홍광석, "무제한 어휘 독립 단어 인식 시스템의 구현", 음성통신 및 신호처리 학술대회 논문집 제 17권 제1호, pp. 145-148, 2000.

[3] 양원렬, 윤재선, 홍광석, "영한 음차 변환을 이용한 무제한 음성인식 및 합성기의 구현", 한국 신호처리·시스템학회 논문집 제1권 제1호, pp181-184, 2000.

[4] 윤재선, "한국어 음성인식 Dictation System 구현" 박사학위논문, 2001.

[5] 윤재선, 홍광석, " VCCV 단위를 이용한 어휘독립 음성인식 시스템의 구현", 한국음향학회지 21권 2호, pp160-166, 2002.

[6] 김형순, 김회동, 이병근, "음성처리기술의 현황과 전망," 한국통신학회, 제8권 제 6호, 1991

[7] L.Rabiner and G.H.Juang, "Fundamentals of Speech Recognition," Prentice Hall, 1993.