

합성음성평가를 위한 다음절 무의미단어 생성과 이용에 관한 연구

A Study on the Generation of Multi-syllable Nonsense Wordset for the Assessment of Synthetic Speech

조철우*, 김경태**, 이용주***

(Cheol-Woo Jo*, Kyung-Tae Kim**, Yong-Ju Lee***)

요 약

인간과 기계의 가장 자연스러운 의사소통의 형태인 음성을 통한 인터페이스를 위하여 여러가지 음성합성, 인식기법들이 제안되고 실용화되고 있다. 특히 음성합성의 경우는 실용화가 상당히 이루어지고 있음에도 불구하고 평가기법에 관하여는 아직도 초보적인 단계에 머물고 있다. 본 논문에서는 무의미 단어에 의한 합성음 평가법에 사용할 수 있는 다음절 무의미 단어군 작성법을 제안하고 실제로 구현되어 있는 규칙합성기를 제안된 단어군에 의해 평가한 사례를 소개하고자 한다. 제안된 단어군 작성방식은 음소단위 명료도 및 음소환경에 관한 평가를 행할 경우 유용하게 사용될 수 있다.

ABSTRACT

These times many kinds of man-machine Interfaces using speech signal, speech recognizers or speech synthesizers, are proposed and utilized in practice. Especially speech synthesis system is widely used in our life. But its assessment method is still in its first stage. In this paper we propose a method to generate multi-syllable nonsense wordset for the purpose of synthetic speech assessment and applies the wordset to one commercial text-to-speech system. Some results about the experiment is suggested and it is verified that the method to generate a nonsense wordset can be used to assess the intelligibility of the synthesizer in phoneme level or in phonemic environmental level.

1. 서 론

음성합성의 최종적인 단계는 규칙음성합성 시스템 또는 문자로부터 음성을 합성할 수 있는 시스템이라고 볼 수 있다. 우리말의 규칙합성에 관한 연구결과가 여러가지 문헌을 통하여 제시되고 있고 상용 시스템까지 등장한 상태이나 이러한 시스템들이 어느 정

도로 명료도와 자연성을 갖고 있는지 평가가 된 경우는 거의 없는 실정이며 평가를 수행한 경우도 단지 단어나 음절수준의 인지도를 측정할 경우가 대부분으로 합성기 개발과 성능향상을 위한 진단평가나 합성기의 선배를 위한 비교 척도로서의 평가는 아직 제대로 이루어지지 못하고 있으며 지금까지 발표된 합성기에 관한 소수의 평가마저도 과정이나 사용 데이터베이스가 알려져 있지 않다.

이는 국내의 여건상 지금까지 합성기자체의 개발에 치우친 점이 없지않아 제대로 된 평가를 행할 여유 없었기도 하거니와 합성음성 평가에 관한 자료를 쉽게 구할수 없었던 사정도 있다. 그러나 음성합

* 창원대학교 제어계측공학과
 ** 한남대학교 정보통신공학과
 *** 전자통신연구소 자동통역연구실
 현재 원광대학교 전자계산기공학과
 접수일자: 1994년 3월 31일

성기의 개발 기술과 합성을 위한 음성분석기술이 어느정도 축적된 지금은 개발자의 입장에서 볼 때 합성음의 명료도와 자연성의 개선을 위하여 보다 손쉽게 행할 수 있는 체계적인 방법이 요구되고 있으며, 차후 합성기의 수요가 다양한 분야로 늘어날 것에 대비하여 구매자의 입장에서 각 합성기간의 성능을 비교해 볼 수 있는 척도가 필요한 실정이다.

이에 본 논문에서는 합성음평가에 사용할 수 있는 다음절 무의미 단어군을 발생시키는 방법과 실제 합성기를 대상으로 평가한 결과에 관하여 논하였다.

II. 합성음평가의 중요성과 현황

외국의 경우는 DECTalk 등과 같은 뛰어난 성능의 합성기 개발 과정에서 여러가지 평가방법이 사용되어 왔고 이를 기준으로 현재 개발중인 합성기의 성능 평가도 이루어지고 있다. EC의 경우 Esprit 계획의 하나로 음성 인식과 합성시스템 평가기술에 대한 대규모의 공동연구가 여러 국가간에 공동으로 이루어지고 있다.(SAM project: Speech Assessment Method project)[1][2][3] 그리고 일본어에 관하여도 몇가지 사례가 보고된 바가 있다.[4][5][6] 또한 평가법의 표준화에 관한 시도도 활발히 행해지고 있다.[7][8][9][10] 그러나 한국어에 관하여는 최근들어 평가에 관한 관심이 고조되고 있고 관련 논문들이 등장하고 있기는 하나 대부분 기존 단어군을 이용하거나 다음절을 이용한 명료도 평가에 머물고 있다.[11][12][13]

이에 효과적으로 우리말에 대한 합성음평가를 할 수 있게 하는 기준과 도구, 데이터베이스의 마련을 위해 본 연구를 시작하였다. 이를 위하여 저자들은 이미 한국어에 대한 합성음성평가의 필요성[14], 현황과 무의미단어에 의한 평가를 활용하는 방안에 관하여 발표한 바가 있다.[15]

본 논문에서는 음소환경을 고려한 합성음성평가용 다음절 무의미단어군의 작성방법과 이를 이용한 평가시험에 관하여 기술한다.

III. 합성음평가를 위한 다음절 무의미단어 생성

합성음성의 평가를 위하여는 기존의 MRT(Modified Rhyme Test), DRT(Diagnostic Rhyme Test) 방법외에 무의미 단어를 이용한 방법이 사용되어 왔

는데, 무의미단어를 이용한 평가법은 합성음의 명료도 평가를 위하여 제안되어 사용되어 왔다.[1][2][3] 이러한 무의미단어에 의한 평가법을 사용할 경우의 장단점은 참고문헌[15]에 잘 나와 있다.

본 논문에서는 지금까지 합성음성의 평가에 많이 사용되어온 MRT, DRT의 장·단점을 고려하여 무의미단어에 의한 합성음성 평가법에서의 장단점, 문제점을 비교, 연구하고, 무의미단어의 작성과정, 제작과정 등을 연구하여 실제의 합성음성평가에 사용하였다.

특히 본 연구에서는 음소환경을 고려한 다음절 무의미단어를 컴퓨터 프로그램에 의해 자동작성하는 방법에 관하여 연구하였다. 유럽의 SAM 계획에서 수행된 것과 같은 기존의 무의미단어 발생기에 관한 연구에서는 주로 단일 음절 무의미단어를 대상으로 명료도를 측정하여왔다. 독일어, 불어, 이탈리아어 등에서 사용된 이와 같은 방법들은 통계적으로 구해진 음소나 음소환경의 출현 빈도수를 바탕으로 시험에 사용한 무의미단어를 발생시켜 평가에 사용하는데, 평가에 사용되는 음소환경의 수가 합성기음성에서 발생할 수 있는 여러 음소환경의 경우를 다루기에 충분하지 못하다. 그리고 많은 환경을 포함하기 위해서는 단어의 수가 많아지게되어 평가시험을 수행하는데 피험자의 협조성 상실등의 현실적인 어려움을 초래할 수 있다. 이와 같은 어려움을 개선하기 위하여 본 연구에서는 자연어와 유사한 음소환경의 비율 갖는 다음절 무의미단어를 자동 작성하는 방법에 관하여 제안하고 연구하였다. 다음절 무의미 단어는 같은 수의 무의미 다음절에 비해 많은 종류의 음소와 음소환경을 시험할 수 있다는 것이 장점이다. 그림 1은 다음절 무의미 단어를 작성하는 흐름도이다.

다음절 무의미단어를 작성하는 방법은 다음과 같이 두가지로 접근할 수 있다. 먼저 음절수를 제한하고 음소환경의 가지수를 택하는 방법이 있다. 이 방법은 자연단어의 평균음절수가 3-4음절이고 음절수가 많을 경우 효과적인 시험이 어려워질 것이라는 가정하에 택하는 방법이다. 그 다음 음절수를 제한하지 않고 모든 음소환경의 종류수가 전체적으로 일정 비율에 달할 때까지 단어를 발생시키는 방법이다. 후자의 경우에도 시험의 효율성을 위해 50-150개 정도로 단어군을 만드는 것이 바람직하다고 생각된다. 본 연구에서는 전자의 방법을 택했다. 그리고 각 음소환경

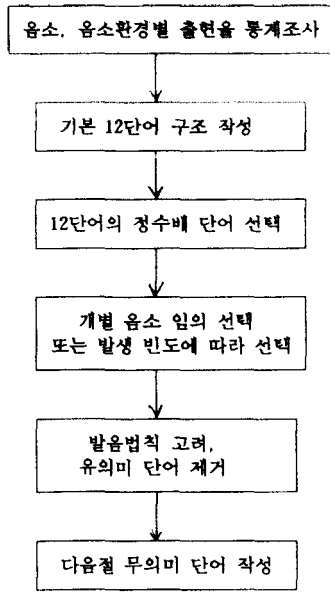


그림 1. 다음절 무의미단어 생성 흐름도

의 수를 모두 포함하더라도 수백개 이상이 되면 실제 실험에의 적용시 효율성이 떨어지므로 단어수를 제한하고 환경수의 비를 자연언어군과 유사하게 하는 방법을 사용하였다. 그러므로 본 단어군은 개개 음소의 정오율을 판별하는 것보다는 오류가 많은 음소환경의 종류를 판별하는데 사용할 목적으로 작성되었다.

단어군작성은 한국어 발음사전 통계조사로부터 가장 빈도수가 높게 나타난 음소환경인 CVC, VCV와 CC, VV환경을 중심으로 행하였다. 단어군 작성의 기준으로는 각 자모음소의 통계적 빈도수를 고려하고, CVC, VCV의 순으로 여러가지 환경이 동일 빈도로 출현하도록 자모음소를 추가하면서 각 음소환경의 출현빈도를 측정하는 방법을 사용하였다.

기본적으로는 각 환경이 1회씩 출현하는 것을 목표로하였고 음소의 빈도수는 고려하지 않았다. 본 연구에서는 혼동의 가능성이 적은 6개의 모음과 18개의 초성자음 종성자음을 결합한 48개로 단어의 갯수를 한정하고 모든 단어가 동일한 음절수를 가지도록 단어군을 작성하였다. 48개의 갯수는 음소환경을 고려한 기본 12단어의 4배수를 택한 것으로 50개정도가 실험에 효과적이라고 생각되어 택하였다. 작성된 단어군으로부터 합성된 음성을 이용하여 한개의 상용

가치음성합성시스템을 시험적으로 평가하였다.

IV. 평가실험 및 검토

청취실험에는 13명의 정상 청각을 가진 20대-30대 남성을 대상으로 행하였으며 시험전에 몇몇 합성음을 시험적으로 들려주었다. 시험은 어학실습실에서 헤드폰으로 청취하도록 하였으며 주어진 답안지들은 합성음을 받아쓰게하는 방법을 사용하였다. 청취자중 2명은 동일한 무의미 단어를 자연음성으로 발성하여 녹음한 것을 들려주었다.

실험에 사용된 합성음은 합성기에서 합성된 음성 Sony사의 DAT에 녹음하여 어학실습실의 방송시설에 연결하여 들려주었다. 청취실험에 앞서 각 청취부스의 청음 수준을 테스트하여 적절한 크기로 청취되는지를 확인하였다. 실험과정은 우선 실험에 사용될 음성이 무의미 단어임을 미리 알려주고 각 음성을 2회씩 들려주었다. 청취자는 들은 음성을 배부된 답안지에 받아쓰도록 하였다.

이 실험에서 사용된 단어군과 포함된 음소환경의 종류는 표1.과 같다. 이 표에 제시된 단어군은 본 연구에서 개발한 무의미 단어 생성기로 작성한 단어군의 한가지 예이며 실제로는 유사한 출현비를 갖는 서로 다른 무의미 단어군을 작성할 수 있다. 표2.는 음소환경별, 대상자별 정답율을 나타낸 것이며 표4.는 단어군에 출현한 음소의 빈도를 보인 것이다.

실험결과 음소환경별 정답율은 무의미 단어에서는 52.75.0%이고 자연음성에서는 24.187.5%를 나타내었다. 또한 자연음성이 VV, VCV, CVC, CC의 순으로 환경별 인식율이 높게 나타난 데 비해 합성음의 경우는 자연음성보다는 낮지만 VV, VCV, CC, CVC의 순으로 인식율이 높게 나타나서 제안된 단어세트가 자연음성과 유사한 인자특성을 갖는다고 볼 수 있다. 그러나 표2.에서 볼 수 있듯이 인식율에 대한 개인차가 심한 경우가 있는데, 예를 들어 청취자 C, G의 경우는 다른 청취자에 비해 모든 음소환경에 대해 낮은 인식율을 보이고 있다. 이것은 다른 화자와는 달리 이러한 특정화자에 대해 유의미 단어에 관한 선입관등의 요인으로 인해 실험이 잘 이루어지지 않은 경우로 분석할 수 있으므로 최종 통계에서 제외하였다. 그리고 4가지 음소환경에 대해 자연음성의 경우와 합성음성의 경우를 비교해 보았을 때 평균 인식율의 비가 VV 38.4%, CC 38.27%, VCV 28.6%, CVC

표 1. 단어말 음소환경의 수의 일례

	CVC	VCV	CC	VV
수무어출...	슴, 출	우무, 어크	0	우어
서소쓰빨...	섯, 숲, 빨, 쓰	어소, 오쓰	0	0
븐스파어...	븐, 슌	으파		아어
쓰입쑤머...	쑤	우머	ㅂㅅ	으이
아카복븐...	복, 븐, 칼	아카, 아로	ㄱㅁ	0
이싸거카...	싸, 겐	이싸, 아거, 어카	0	0
아우땃금...	땃, 금	우땃	ㅅㄱ	아우
오언쓰세...	쓰	으세	ㄴㅅ	아우
어더탄땃...	덜, 탄, 땃	어더, 어타	ㄴㅂ	0
플켓콜루...	플, 켓, 콜	0	ㄹㄱ, ㅅㄱ, ㄹㄴ	0
물뚝속쏘...	물, 뚝, 속	0	ㄹㄷ, ㄱㅅ, ㄱㅁ	0
관쪽닷저...	관, 쪽, 닷	0	ㄷㅁ, ㄱㅅ, ㅅㅁ	0
버키엘트...	덜, 키에	어키,	ㄷㅅ	이어
시씨비땃...	섯, 씌, 비땃, 땃	어씨, 어비, 어땃	0	0
츄러티어...	츄, 띵	어티,	ㄴㄹ	이어
렘도쳐코...	렘, 돛, 칠	오쳐, 어포	ㅁㅅ	0
오루븐족...	루, 븐, 족	오루, 우븐	ㄴㅎ	
오로쑤카...	룻, 쑤	오로, 오쑤 우카	0	0
오아츄니...	츄, 니	아츄,	ㄱㄴ	오아
아울허부...	허	어부	ㅅㅎ	아으
어싸쑤땃...	섯, 쑤, 땃	어싸, 아쏘	ㅂㅅ	0
힌켄탁히...	힌, 겐, 탁	0	ㄷㄱ, ㄷㅅ, ㄱㅎ	0
짱국쑤까...	짱, 국, 쑤	0	ㄱㅁ, ㄷㄱ	0
жат문듭주...	жат, 문, 듭	0	ㅅㅁ, ㄴㅁ, ㅁㅅ	0
기버러쳐...	길, 덜, 렷	어터, 어러 어쳐	0	0
타크더흔...	탈, 큰, 덜	아크, 으더 어흔	0	0

꽃파치우...	꽃, 팟	아치	ㅅㅁ	아우
줄커로스...	줄, 켈, 롯	0	ㅅㅁ	0
오쳐킵쑤...	쑤, 켄, 쑤	오쳐, 어끼	ㅂㅅ	0
어무차마...	물, 참	어무, 아마 우차	0	0
으오쪽갓...	쪽, 갓	오쑤	ㄱㅁ	으오
우이코찌...	꼴	이코, 오찌	0	우어
오쁘뚝뚝...	쁘, 뚝, 뚝	오쁘, 오뚝	ㅅㅅ	0
뚝켄잡쑤...	뚝, 겐, 잡	0	ㅂㅅ, ㄷㅅ, ㅂㅁ	0
롬싼쑤꺼...	롬, 산, 쑤	0	ㅂㅁ, ㄴㅁ, ㅅㄱ	0
익롬짜브...	익, 롬, 짜	0	ㄱㅅ, ㅂㅁ, ㄱㅂ	0
히누이츄...	힌, 츄	이누, 이츄	0	우어
쳐호쯔닐...	츄, 호, 쑤, 쑤	어호, 오쯔 으니	0	0
덜꺼지어...	덜, 겐	어지,	ㄷㄱ	이어
죽카르베...	죽, 칼, 르	아르, 으베	ㄱㅁ	0
어쑤뽕땃...	쑤, 뽕, 땃	어쑤, 우쑤	ㅂㅁ	0
이터포로...	덜, 풀	이터, 어포 오로	0	0
어으칙쑤...	칙, 쑤	으쳐,	ㄱㅅ	어으
어익구무...	굶	우무	ㄱㄱ	어어
으짜극날...	극, 날	으짜, 어그	ㄱㄴ	0
짐선랍페...	짐, 선, 략	0	ㅂㅁ, ㅁㅅ, ㄷㅂ	0
쑤순죤씨...	쑤, 순, 죤	0	ㅂㅁ, ㄷㅅ, ㄷㅁ	0
뽕땃땃저...	뽕, 땃, 땃	0	ㅅㅁ, ㅅㅅ, ㅅㅅ	0
종류수 계	121	65	58	16

자연어에서의 음소환경의 비율

CVC	VCV	CC	VV
43.8%	22.1%	19.9%	9.9%

는 개략 4 : 2 : 2 : 1 이 된다.

그러므로 본 단어세트는 자연어와 유사한 환경의 비율 가진다고 볼 수 있다.

표 2. 합성음성에 대한 음소환경별 정답율 표

(정답율: %)

과목	CVC	VCV	CC	VV
A	1.90	12.30	16.00	31.00
B	8.00	18.50	15.50	31.30
C	6.60	8.20	5.20	12.50
D	12.45	16.67	12.07	43.75
E	14.05	16.67	10.34	18.75
F	6.61	15.15	1.72	18.75
G	9.09	6.06	6.90	6.25
H	6.61	12.12	10.34	12.50
I	18.18	25.76	13.79	75.00
J	13.22	25.76	8.62	50.00
K	10.74	18.18	12.07	43.75
평균	10.04	15.76	10.23	31.23

표 3. 자연음성에 대한 음소환경별 정답율 표

(정답율: %)

과목	CVC	VCV	CC	VV
X	32.23	43.08	24.14	75.00
Y	42.15	63.64	29.31	87.50
평균	37.19	53.36	26.73	81.25

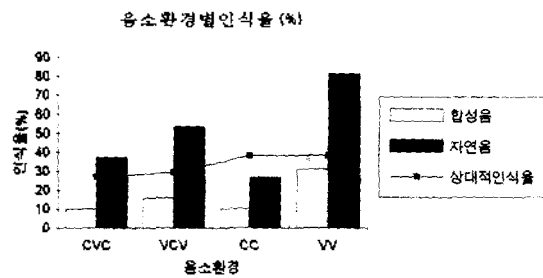


그림 2. 음소환경별 정답율의 비(합성음성/자연음성 * 100%)

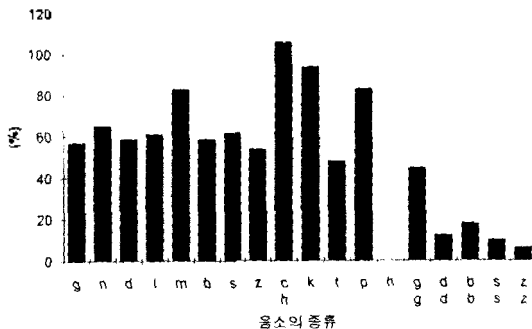


그림 3. 음소별 정답율의 비(합성음 인식율/자연음 인식율 * 100%)

표 4. 단어군에 출현한 각 음소의 빈도수

(초성자음)	(보음)
ㄱ: 8	ㅏ: 27
ㅋ: 11	ㅓ: 40
스: 7	ㅣ: 33
ㅎ: 8	ㅑ: 33
ㅌ: 6	ㅗ: 35
ㄷ: 5	ㅡ: 27
ㅍ: 9	
ㅊ: 7	(중성자음)
ㄴ: 4	ㅇ: 0
ㅍ: 5	ㄹ: 1
ㅌ: 13	ㄱ: 1
ㅍ: 7	ㅇ: 0
ㅊ: 4	ㅍ: 0
ㅆ: 23	ㅌ: 1
ㅃ: 9	ㄷ: 1
ㅆ: 6	
ㅋ: 9	
ㄹ: 10	

26.90%로 CVC의 경우가 가장 자연음성에 비하여 낮은 인식율을 보이고 있으므로 이런 음소 환경에 대한 분석이 더 세밀하게 이루어져야한다는 뜻으로 해석할 수 있다.

그리고 단어군에 등장하는 각 음소에 대한 측정 결과는 표5에서와 같은데 합성음성이 자연음성에 비해 ㄱ, ㄴ, ㄷ, ㄹ, ㅌ, ㅍ, ㅊ, ㅆ 등은 36-58% 정도의 인식율을 보이고 있는데 비해 ㅈ, ㅋ 등은 상대적으로 높은 인식율을 보이고 있다. 그중에서도 초음 합성음의 경우가 더 높은 인식율을 나타내어서 이상하게 생각될지 모르겠으나 이것은 합성음에서 파찰을 나타내어서 이상하게 생각될지 모르겠으나 이것은 합성음에서 파찰음이나 파열음의 무성파열부분을 강조하기 위해 잡음성분으로 강조해준 것을 원인으로 추측할 수 있기 때문에 음성학적으로 충분한 이유가 있는 결과이다. 반면 ㅍ, ㅊ, ㅆ 등의 경우는 1.1-9.4%의 아주 낮은 인식율을 보이고 있는데 이는 경음의 합성하기가 힘든 것을 보여주어 시험대상 합성기에서 경음에 대한 많은 개선이 필요함을 나타내 준다. 그러나 자연음의 경우는 표6에서 보이듯이 상대적으로 높은 인식율을 보였다.

표6과 표7은 각각 자연음성과 합성음성의 경우 단어군에 등장한 음소별 혼돈현상을 표로 만든 것이다. 표6에서 보면 자연음성에서도 경음에서 잘못 인식되는 현상이 두드러지게 나타나고 있음을 알 수 있는데 이는 표7에서 합성음의 경우 가장 큰 오인식을 가져오는 원인이 되고 있다. ㅎ의 경우는 합성음에서 전

표 5. 각 음소별 정답율 % (합성음, 자연음)

	ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅈ	ㅊ	ㅋ	ㆁ	ㆅ	ㅇ	ㅍ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ	
합성음	36	41	53	55	83	44	56	50	77	88	44	58	0	16	9.4	6.9	1.1	5.4						
자연음	63	63	90	90	100	75	91	93	73	94	92	70	75	36	75	39	11	86						

표 6. 음소별 혼돈표 (자연음성)

	ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅈ	ㅊ	ㅋ	ㆁ	ㆅ	ㅇ	ㅍ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ		
ㄱ	10																							16	
ㄴ		5																							5
ㄷ			9																						9
ㄹ		2		18																					21
ㅁ					18																				18
ㅂ						9																			23
ㅅ							20	1																	62
ㅈ								13	4																19
ㅊ									16																16
ㅋ										17															20
ㆁ											24														27
ㆅ												7													7
ㅇ										1															13
ㅍ													12												11
ㅑ	5												1	5											8
ㅓ											2														10
ㅕ						3																			7
ㅗ							2																		15
ㅛ									2																12
ㅜ																									5
ㅠ																									15
ㅡ																									5
ㅣ																									5
공난	1	1	1	2																					5
합계	16	8	10	20	18	12	22	14	22	18	26	10	16	14	6	18	46	14							312

표 7. 음소별 혼돈표 (합성음성)

	ㄱ	ㄴ	ㄷ	ㄹ	ㅁ	ㅂ	ㅅ	ㅈ	ㅊ	ㅋ	ㆁ	ㆅ	ㅇ	ㅍ	ㅑ	ㅓ	ㅕ	ㅗ	ㅛ	ㅜ	ㅠ	ㅡ	ㅣ		
ㄱ	23				1	2																			73
ㄴ	1	13	3	18	6	1																			43
ㄷ			21	4		2								3											56
ㄹ	6	1	3	44	1																				59
ㅁ	1	5		2	60	8																			88
ㅂ				2		21	1																		68
ㅅ			2				49	5	16	2	2	1	8	2											155
ㅈ	7		3			1	23	28			3	1	10	5	2										193
ㅊ	2						6	6	68																120
ㅋ	14					1					63	27	7	1	6										122
ㆁ			6			2						46	6												65
ㆅ			1			2		1		1	13	23													45
ㅇ								1		2	4														10
ㅍ	4				1	1				1															23
ㅑ																									3
ㅓ						2																			9
ㅕ								2																	6
ㅗ							2	2																	19
ㅛ	1						1	2																	10
ㅜ																									3
ㅠ																									91
ㅡ	5	13	1	9	4	5	6	4	4	1	3	1	12	2											91
ㅣ																									1248
합계	64	32	40	80	72	48	88	56	88	72	104	40	64	56	32	72	184	56							1248

혀 인식이 되지 않고 있는데 이는 단어중간에 놓이
울 경우 약화되거나 생략되도록한 합성규칙때문일
수 있다. 이는 또한 단어의 작성에서 음운변동현상이
많이 고려되지 않은것이 원인일 수도 있으므로 앞으로
이를 검토하여 보완해야할 김이다. 그림2는 자연
음에 비교한 합성음의 인식율의 비율 %로 나타낸 것
이다. 여기서도 역시 성운의 경우가 잘 인식되지 않
고 있음을 알 수 있다.

본 연구에서 작성하고 사용된 단어군은 음소환경
의 발생빈도의 비가 대규모 단어군과 유사하도록 작
성되었으며, 개개 음소의 출현빈도의 비도 거의 유사
하게 만들어져 있어서 음소환경을 고려한 합성음성
시험에 적절한 것으로 보인다. 그리고 결과면에서도
음성학적으로 근거가 있는 오류들을 보여 주고 있으
므로 개개 음소의 명료도 판별에도 유용할 것으로 생
각된다. 또한 자연어와의 인식을 비교에서도 자연어
와 합성음의 결과가 유사한 특성을 보여주고 있으므
로 제안한 무의미 단어가 유효함을 보여준다. 그러나
아직도 다음과 같은 몇가지 보완되어야할 문제점들
이 있으므로 차후의 연구에서 개선될 점으로 생각된
다.

이러한 개선될 점들로는 단어군 작성시의 통상적
발음법칙의 반영, 즉, 자연언어군에서 어떤 음소뒤에
어떤 음소가 확률적으로 많이 나타나고 어떤 음소가
울 확률이 거의 없다 등의 발음법칙의 적용이 필요하
다. 본 연구에서는 그러한 음운결합현상에 의해 나타
나는 차이를 결과 측정시에 수작업으로 고려하여주
었다. 이때 고려해준 법칙은 연음법칙, 말음법칙, 자
음집변동 기본적인 음운변동법칙들이다. 그리고 음
소열이 겹칠 때 발생하는 음운변동에서 새로 파생되
는 음소환경을 고려하는 것이 필요하다. 또한 현재
포함되지 않은 소수 환경의 종류를 포함하는 것도 필
요하며, 앞으로의 청취실험에서는 동일한 단어표에
대해 자연음성을 같이 섞어서 실험할 필요도 있다.

그리고 본 연구에서 제시된 단어의 음절수가 4개가
되므로 변화가 크지는 않지만 운율의 영향을 받게되
어 인지도에도 영향을 미칠 수도 있다. 현재는 합성
기 내부의 특성을 제어할 수 없는 형편이므로 합성기
가 갖고 있는 운율발생정보에 따라 합성음을 만들고
있으나 앞으로 가능하다면 제어된 환경을 부여해 주
는 것도 생각해 볼 수 있겠다. 그래서 어양, 길이들의
운율환경을 단순화 한다면 명료도시험만의 목적으로

제한하여 본 연구에서 생성한 단어군을 사용하는 것
이 유효하다고 생각된다. 그러나 4음절 무의미 단어
의 경우 개개 동일한 운율 정보가 부가되게 되므로
운율정보의 변화에 따른 명료도 인식에의 변동 요인
을 가지고 생각된다. 본 실험에서 사용한 합성기의
경우는 끝음절에서 기본 주파수가 하강하는 현상을
보였고 단어 중간에서는 유사한 음운 정보를 가졌다.

그러나 모든 조건을 갖춘 단어군을 만드는 것은 매
우 복잡한 문제가 되므로 실제 단어군에 대한 통계학
적 연구와 함께 앞으로도 계속적으로 연구되어야만
할 문제이고, 이를 위해 음성학, 국어학, 음성공학자
들이 계속 연구해야할 것으로 생각된다. 이와 같은
문제점들 때문에 본 연구에서와 같이 일정한 제한조
건하에서 단어군을 만들어가는 접근방법이 유효할
것으로 생각된다.

V. 결 론

본 논문에서는 합성음성을 평가하기 위한 다음질
무의미단어의 생성법을 제안하고, 이 방법을 이용한
합성음성의 평가사례를 제시하였다.

민저 무의미단어가 가져야할 조건들을 제한하고,
그 범위내에서 한국어 음소의 통계치를 참조하여 무
의미 단어를 발생시킬 수 있는 소프트웨어를 작성하
였고, 제안된 방법에 의해 48개의 4음절 무의미 단어
군을 발생시켰다. 작성된 무의미 단어군을 이용하여
13명의 청취자를 대상으로 어학실험실의 방음부스에
서 청취실험을 행하였다. 실험결과를 분석한 결과 음
소환경의 비율면에서 자연음성과 유사한 특성을 가
지며 완전하지는 않지만 여러 음소를 일정하게 빠짐
없이 포함하고 있고, 측정된 혼돈현상이 음성학적이
로 근거가 있으므로 앞으로의 합성음평가 및 합성기
의 개발과 성능개선을 위한 합성음성평가에 사용할
수 있는 방법으로 생각된다.

아울러 앞으로 앞의 검토사항들에서 지적한 내용
을 보완하고 단어군 발생과 시험 및 평가를 자동화
할 수 있는 소프트웨어개발을 차후 계속적으로 수행
할 예정이다.

*본 연구는 전자통신연구소의 음성입출력시스템의 성능
평가법 연구의 일부로 수행되었습니다.

참 고 문 헌

1. Multi lingual Speech Input/Output Assessment, Methodology and Standardisation. Final Report, SAM-UCL-G004, 1992.
2. Speech Input/Output Assessment and Speech Database. Proceedings of ESCA workshop, 1989.
3. Louis C.W. Pols, "Evaluating the Performance of Speech Technology Systems," Institute of Phonetic Sciences, University of Amsterdam, Proceedings 15, 27-41, 1991.
4. 北脇信彦, "무호화, 합성음질의 평가," 일본 전자정보통신학회지, 70권, 4호, 1987.
5. 比企静雄 의, "음성처리기술의 성능평가법에 관한 요인," 문부성 과학연구비 음성언어에 따른 맨머신 인터페이스의 고도화 연구보고서, PASL 63-8-1, 1988.
6. 粕谷英樹, "음성합성기술의 평가," 문부성 과학연구비 중점영역 연구 음성언어에 따른 맨머신 인터페이스의 고도화 연구보고서, PAS : 62-8-1, 1988.
7. "Experiment in assessing the quality of synthetic speech," Temporary Document No.70-E, CCITT working party 5/XII, Geneva, 27 Feb.-3 March 1992.
8. "Elements for Draft Rec. on synthetic speech assessment," Temporary document 80-E, CCITT working party 5/XII, Geneva, 27 Feb.-3 March 1992.
9. "Report on Question 5/XII, Speech synthesis/recognition systems," Temporary Document No.52(Rev. 1)-E, CCITT working party 5/XII, Geneva, 11-19 May 1993.
10. "Draft Recommendation P.8S-Subjective Performance Assessment of the Quality of Speech Voice Output Devices," COM 12-6-E, CCITT working party 5/XII, 1993.
11. 김정환, 장대영, 강성훈, "병료도 평가용 단음절 목록의 개발과 타당성 검증," 한국음향학회 학술논문발표회 논문집, 제12권, 1호, 1993.
12. 강찬희 외 5인, "음절유형별 규칙합성을 음질평가," 한국음향학회 학술논문발표회 논문집, 제12권, 1호, 1993.
13. 정유현, 최준혁, 한민수, "Text-to-Speech 합성음 품질평가," 한국음향학회 학술논문발표회 논문집, 제12권, 1호, 1993.
14. 이용주, 김경태, "음성인식 및 합성의 성능평가법에 관하여," 제5회 신호처리 합동학술대회 논문집 제5권 1호, 1992.
15. 조철우, 김경태, 이용주, "무의미 단어에 의한 규칙합성음의 평가 및 진단법에 관하여," 제10회 음성통신 및 신호처리 워크샵 논문집, SCAS-10권 1호, 1993.

▲조 철 우(Cheol-Woo Jo) 1961년 9월 12일생



1983년 2월 : 고려대학교 공과대학 전자공학과 졸업(공학사)

1985년 2월 : 고려대학교 대학원 전자공학과 졸업(공학석사)

1989년 2월 : 고려대학교 대학원 전자공학과 졸업(공학박사)

1985년 8월 ~ 1986년 11월 : 전자통신연구소 음향연구소(위촉연구원)

1992년 2월 ~ 1993년 2월 : 영국 Keele University Department of Communication and Neuroscience (Visiting Research Fellow)

1989년 3월 ~ 현재 : 국립 창원대학교 제어계측공학과 조교수

소속학회 : 한국음향학회 종신회원, 대한전자공학회 종신회원, IEEE 정회원, 영국 음향학회 MIOA

※주관심분야 : 음성 및 음향신호의 모델링, 음성의 규칙합성, 음향신호의 재추 및 합성, 합성음성의 평가, 그외 음성 및 음향신호처리에 관심있음

▲김 경 태(Kyung-Tae Kim) : 제 12권 1호 참조

▲이 용 주(Yong-Ju Lee) : 제 12권 1호 참조