음절 유형별 규칙합성음 음질평가

姜 贊 熙*°, 李 宗 憲**, 權 奇 衡**,
安 定 根**, 徐 成 泰**, 陳 庸 玉**

*, 상지대학교 병설 전문대학 전자과 **,경희대학교 전자공학과

(The Evaluation of Speech Quality Synthesized by Rule According to Korean Syllable Types)

(Chan Hee Kang*°, Jong Heon Lee**, Ki Hyung Kwon**,
Jeong Keun An**, Sung Tae Sea**, and Yong Ohk Chin**)

*, Dept. of Electronics in Sangji junior colledge

**, Dept of Electronic Engineering in Kyunghee Univ.

要 約

본 논문은 한국어 문어변환(TTS:Text-to-Speech) 시스템에서의 음성합성시 음질 및 자연성 개선을 위한 연구 결과이다. 합성음 평가방법으로는 한국어 발음대사전에 수록된 빈도수 순위대로 추출한 음절(V형: 19개, CV형:80개, VC형:30개, CVC형:100개, 총 229개)을 대상으로 규칙합성시킨 1음절어(합성음절수:229개)중 음절유형별로 15개씩 총 60개 음절을 20초간 3 회 반복음의 녹음 테이프를 작성한 합성음에 대하여 사전지식이 없는 임의의 그룹을 선정하여 이해도, 명료도, 잠음감, 자연성등 4 가지 항목에 대하여 오피니온 평가를 수행한 결과를 제시하였다.

I .서 론

언어는 인간과의 의사전달 수단으로써 발성기관을 통하여 생성되어진 음성을 매체로 사용되는 청각적인 정보전달 수단의 한 방편이며, 문자는 시각적인 의사전달 수단이다. 따라서 음성합성의 최소한의 목적은 정보전달에 있으며, 음성합성시 정보전달의 능률을 극대화시키기 위하여는 반드시 자연성을 고려하여야 한다. 음성합성에서의 자연성이란 인간의 다양한 감정이 음성에 표출되어 전달되므로 강약과 장단과 쉼과 억양등과 같은 운율요소를 인위적인 조작으로 합성음에 부여시켰을 때 원음에 일치한 정도라 할 수 있다.

시간영역에서의 운율제어의 어려움이란 원 파형에 임의의 함수를 가하거나 조작 변경시키면 피치주기성을 상실하기 쉬울 뿐만아니라, 합성하고자* 하는 파형이 자연 생성되어진 파형보다 심하게 왜곡되어 음질이 저하되거나 변질되는데 문제가 있다. 본 논문에서는 이러한 시간영역에서의 합성방식이 지니고 있는 운율제어의 한계성을 극복하여 양질의 규칙합성음을 발생시키고자 하였다.

II. 음성합성 알고리즘

시간영역에서의 규칙합성시 텍스트로 부터 변환된 음소 기호열에 따라서 저장된 음성 데이타를 악세스하여 합성시킬 경우에는 운율요소의 제어가 용이하지 못하여 서문에서와 같이 자연스런 합성음을 발생시킬 수 없다. 본 논문에서는 이와같은 문제점을 해결하기 위하여 표1에 제시된 파형분석과정을 거쳐 합성시키고자 하는 파형의 진폭, 지속시간과 피치주기간격등과 같은 성분을 합성용 매개변수로 추출시켜 규칙합성용 데이타 포맷 사전을 구성(표2)하였으며, 이를 이용한 합성 결과는 4장에 제시하였다. 음성파형으로 부터 규칙합성용 매개변수를 추출하기 위하여 임의의 음성 데이타를 $x(n)$, 단음절의 데이타 갯수를 N, 단음절내에서의 1 피치주기의 프레임 갯수를 N_P로 각각 정의하면 단음절의 음성 데이타 열은 $\sum_{n=1}^{N} x(n)$ 으로 표기된다. 이 때 각각의 피치 프레임 구간의 경계를 P_{s1}, P_{s2}, P_{s3}, ···· 등으로 나타내고, 각 피치 프레임 구간에서의 데이타 갯수를 N_{Ps1}, N_{Ps2}, N_{Ps3}, ···· 등을 배열 $N_{Ps}(\ )$로 표기하면, N 개의 음성 데이타 열 $\sum_{n=1}^{N} x(n)$을 1 차원 배열인 1 피치 프레임 단위의 N_P 개 소블록의 합으로 표기 가능하므로 이를 2차원 배열로 표시하면,

$$즉, \quad \sum_{n=1}^{N} x(n) = \sum_{n1}^{Np} \left[\sum_{n2=1}^{N_{ps}(n1)} x(n1,n2)\right] \qquad \cdots\cdots (1)$$

(단, $n=n1+\sum_{n1}(n1-1)\cdot N_{ps}(n1-1)$, $N_{ps(0)}=0$ 임.)

와 같다. 여기서, x(5,10)은 5번째 단위피치 구간의 10번째 데이타를 의미한다. 즉, x(n1,n2)는 x(음절내 단위피치 구간 번호,단위피치 구간내 데이타 번호)를 의미한다. 또한, 단위 피치 프레임 구간내에서의 음성 데이타 열의 최대 진폭의 절대치를 각각 A_{m1}, A_{m2}, A_{m3}, ···· 등으로 정의하고, 각각의 피치 프레임 단위 구간내의 데이타 열을 일정한 크기로 정규화시킨 임

의의 음성 데이타를 $x_N(n)$로 정의하면, 2 차원 분록화 배열로 표시된 음성 데이타 $\sum_{n1=1}^{Np}\sum_{n2=1}^{Npst(i)} x(n1,n2)$은

$$\sum_{n1=1}^{Np}\sum_{n2=1}^{Npst(i)} x(n1,n2) \approx \sum_{n1=1}^{Np}\sum_{n2=1}^{Npst(n1)} A_m(n1)\cdot x_N(n1,n2) \quad \cdots\cdots\cdots (2)$$

로 표시된다. 윗 식들로 부터 추출·저장하여 데이타 포맷 사전에 작성된 매개변수는 총 5개로써 이를 정리하여 보면, 식(1)에서 추출된 매개변수는 1)단음절내 전체 데이타 갯수 정보 N(2 바이트), 단음절내 단위피치 경계검출에 의하여 추정된 2)단음절내 피치 갯수정보 N_P(1 바이트), 3)각 피치 프레임 구간에서의 데이타 갯수 정보 $\sum_{i=1}^{Np}N_{pst}$(N_P 바이트)와 식(2)에서 추출된 4)단위피치 구간내 에서의 최대진폭 정보 $\sum_{i=1}^{Np}Am(i)$ (N_P바이트)와 5)단위피치별로 정규화된 음성데이타 $\sum_{n1=1}^{Np}\sum_{n2=1}^{Npst(n1)} x_N(n1,n2)$ (N바이트)등 이다.

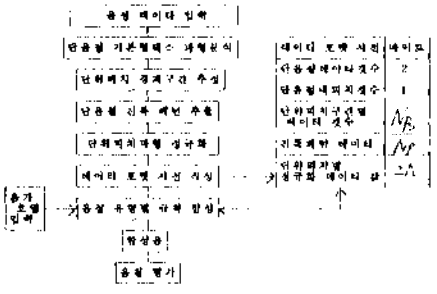

그림 1. 규칙합성 블록도
fig 1. Block diagram of synthesis by rule

Ⅲ: 음성합성예

그림 1은 본 논문에서 사용한 음성합성의 개략적인 과정을 블록도로 표시한 것이다. 이 그림에서와 같이 단음절 단위의 파형이 입력되면 맨 먼저 파형분석 과정을 거쳐 합성에 필요한 파형정보를 추출하는 과정을 거친다. 그림 2에 표시된 CVC형 파형"공"을 한 예로 놀어 설명하면 3장의 식(1)에서 1음절어의 전체 음성데이타는 2,434 샘플 포인트이다. 이를 1 피치주기간격으로 분할하여 표시하면 표 1에서와 같이 23개의 구간으로 분할하여 표시할 수 있으며, 표2의 19번째 규칙합성용 포맷표에 표시된 1 음절내에 존재하는 1 피치주기의 구간갯수 NP는 23으로 주어진다. 이러한 피치구간을 검출하여 각각의 1피치주기를 구하기 위하여는 피치간의 경계를 검색하여 합성에 이용하였다. 이때 각각의 피치주기를 정확히 검출하지 못하면 단음과 장음합성시 2가지의 중요한 잡음이 발생된다. 그림 3은 그림 2의 파형을 장음으로 규칙합성시켰을

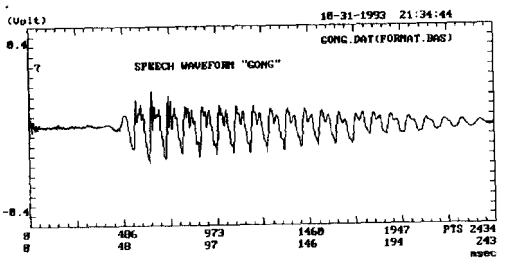

그림 2. 음성파형도 "공"
fig 2. speech waveform of "gong"

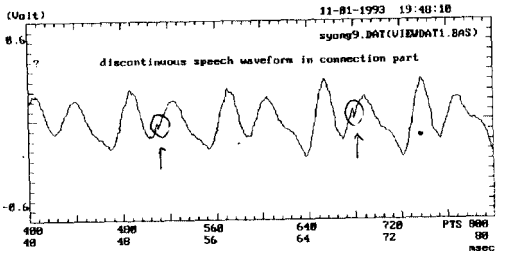

그림 3. 접합면에서의 불연속 잡음
fig 3. discontinuity in connection part

때 이웃간의 파형의 접합부에서 불연속점이 발생하여 찌글거리는 잡음을 유발시키는 한 예를 표시한 것이다. 또 한가지의 잡음은 위상왜곡으로써 추정된 피치주기간격이 짧아지거나 길어지면 잡음이나 단음 합성시 원래의 파형이 지니고 있는 주기성이 흐트러지고 이로 인하여 위상왜곡이 발생하는 요인이 된다. 따라서 1 피치주기의 경계구간을 검출하기 위하여는 상당한 주의를 요하여야 한다. 표1에서 좌측 DATA POINT항은 파형상단부에서의 그 피치주기열 내의 최대값의 위치을 검색하여 표시한 것이며, 우측항의 DATA POINT항은 파형 하단부에서의 최대값 위치를 표시한 것이다. MAX와 MIN항은 최대값과 최소값의 크기를 구한 것으로써 규칙합성시 진폭패턴을 제어할 수 있도록 1 음절이내에서의 최대값에 대한 상대비로 변환하여 표2의 데이타 포맷표(AMP RATIO항)에 저장하여 규칙합성에 이용하였다. 표1에서의 INTERVAL항과 표2에서의 1PERIOD항은 각각의 1피치주기열의 데이타 수를 추정하여 표기한 것이며, 표2에서의 D$은 무성자음이 초성구간에 존재하는 유무를 표시한 것이다. 그리고 표2에서의 NC는 자음소의 데이타 갯수를 표시한 것이며, NTOTAL은 전체 데이타 갯수를 표시한 것이다. MAX는 1음절 데이타중 최대진폭을 나타낸 것으로써 이를 이용하여 강약음의 정도를 규칙합성시키기 위한 파라메타로 사용하였다. 3장의 식(2)에 표시된 우측항의 A_m은 표2에서의 AMP RATIO항에 표시된 값을 나타내며, N_{ps}는 1PERIOD항에 표시된 1피치주기열내의 데이타 갯수를 의미하며,

표 1. 파형분석표
table 1. waveform analysis table

SPEECH SIGNAL (GONG .DAT) ANALYSIS
(FORMAT.BAS)

No.	DATA POINT (PTS)	MAX (mV)	INTERVAL (PTS)	DATA POINT (PTS)	MIN (mV)	INTERVAL (PTS)	No. ZERO CROSSING	MAX RATE	MIN RATE
1	496	56	95	457	-23	70	1	0.30	-0.12
2	554	132	70	543	-118	85	2	0.72	-0.63
3	635	171	83	626	-147	84	4	0.93	-1.00
4	721	151	88	711	-165	86	6	0.81	-0.88
5	810	97	90	798	-139	88	4	0.52	-0.74
6	901	96	90	887	-158	88	3	0.51	-0.84
7	990	103	89	975	-145	88	4	0.55	-0.78
8	1080	101	90	1063	-141	88	2	0.55	-0.75
9	1169	99	89	1151	-138	89	2	0.53	-0.74
10	1258	85	89	1240	-134	89	2	0.46	-0.71
11	1348	82	90	1330	-127	90	2	0.44	-0.68
12	1437	82	89	1419	-114	89	2	0.44	-0.61
13	1527	75	90	1509	-106	90	2	0.40	-0.56
14	1617	63	108	1599	-99	90	2	0.34	-0.53
15	1743	57	91	1688	-98	90	1	0.31	-0.52
16	1798	58	73	1779	-54	91	2	0.31	-0.29
17	1889	50	94	1870	-49	87	2	0.27	-0.26
18	1986	39	95	1953	-47	85	2	0.21	-0.25
19	2080	30	93	2041	-42	91	2	0.16	-0.22
20	2172	27	93	2134	-31	92	2	0.15	-0.17
21	2265	24	94	2225	-26	90	2	0.13	-0.14
22	2360	17	50	2315	-18	93	1	0.09	-0.09
23	2366	15	71	2421	-10	71	2	0.08	-0.05

표 2. 데이타 포맷 예
table 2. examples of data format

No.	FORMAT NAME	BS	BC	TOTAL	SP	MAX	1 PERIOD				AMP. RATIO				
1	1.FRM	0	0	2099	22	912	128	99	...	87	76	.2	.5	...	.2 .3
2	DA.FRM	1	186	1674	16	2304	114	89	...	93	73	.6	.5	...	.1 .1
3	IB.FRM	0	0	2488	32	2348	99	79	...	49	32	.3	.4	...	.1 .0
4	JEONG.FRM	1	454	2592	22	2716	106	80	...	89	49	.3	.8	...	.2 .1
5	L3.FRM	0	0	2000	22	2060	101	85	...	92	82	.1	.2	...	.1 .1
6	WUBOM.FRM	0	0	2796	33	3320	111	86	...	72	66	.2	.6	...	.1 .1
7	DONG.FRM	1	287	2144	21	2672	88	78	...	85	74	.3	.8	...	.1 .1
8	IL.FRM	0	0	2390	30	2464	92	77	...	79	81	.2	.4	...	.1 .1
9	SAMG.FRM	1	500	2540	27	2956	114	79	...	89	75	.3	.8	...	.1 .1
10	GI.FRM	1	497	1884	15	1884	99	87	...	92	77	.6	.7	...	.1 .1
11	SA.FRM	1	680	2350	19	3048	109	84	...	89	80	.6	.8	...	.1 .1
12	JI.FRM	1	530	2142	18	3456	92	85	...	90	85	.5	.5	...	.1 .1
13	AM.FRM	0	0	2298	28	3528	106	80	...	74	63	.2	.6	...	.2 .1
14	JEOB.FRM	1	475	2614	24	3052	123	85	...	84	67	.5	.9	...	.1 .1
15	O.FRM	0	0	1642	17	2176	125	91	...	99	70	.3	.5	...	.1 .1
16	EUN.FRM	0	0	2742	34	3872	109	82	...	78	76	.1	.1	...	.1 .0
17	MUL.FRM	0	0	1920	22	2644	73	80	...	85	70	.1	.3	...	.2 .1
18	DO.FRM	0	0	1654	18	2152	104	88	...	85	74	.2	.4	...	.1 .1
19	GONG.FRM	1	430	2438	23	2440	90	85	...	93	71	.1	.6	...	.1 .1
20	JA.FRM	1	424	2312	19	2328	98	79	...	84	37	.2	.8	...	.1 .0
21	GA.FRM	1	310	2024	19	3088	103	90	...	82	37	.4	.6	...	.1 .0
22	MOB.FRM	0	0	2540	28	3300	117	92	...	91	52	.1	.3	...	.1 .0
23	A.FRM	0	0	1592	18	2488	99	57	...	91	72	.1	.3	...	.1 .1
24	HA.FRM	1	535	2086	18	3468	109	77	...	81	48	.2	.5	...	.1 .0
25	SEOM.FRM	1	694	2592	22	3612	110	81	...	54	17	.5	.9	...	.0 .0
26	EOI.FRM	0	0	2244	24	1512	108	88	...	87	70	.2	.4	...	.2 .1
27	SU.FRM	1	1512	2932	16	2240	91	79	...	90	115	.4	.6	...	.2 .1
28	EOP.FRM	0	0	1090	13	2696	107	84	...	89	55	.3	.2	...	.2 .0
29	EU.FRM	1	329	1862	17	3204	129	84	...	90	42	.5	.6	...	.1 .0
30	IN.FRM	0	0	2478	31	2048	102	80	...	75	67	.2	.4	...	.2 .1

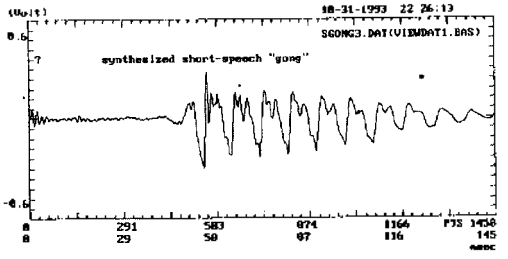

그림 4. CV형 단음 규칙합성 예
fig 4. example of synthesized short - speech by rule

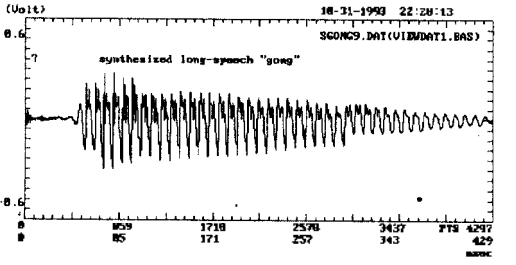

그림 5. CV형 장음 규칙합성음 예
fig 5. example of synthesized long - speech by rule

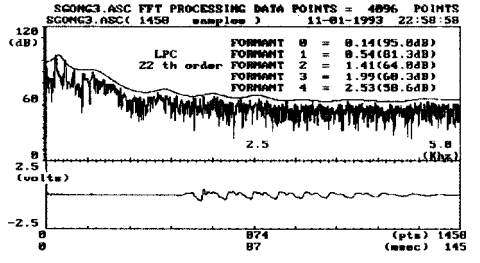

그림 6. 그림4의 포르만트 추정도
fig 6. estimated formant of fig 4.

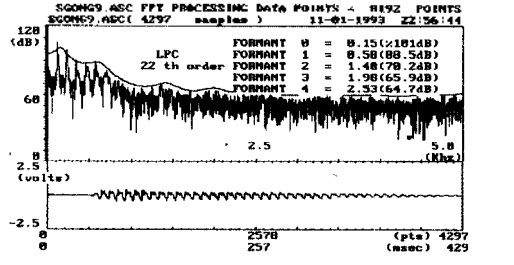

그림 7. 그림 5의 포르만트 추정도
fig 7. estimated formant of fig 5.

N_p는 1음절어내의 단위피치주기 갯수를 표시한 것이다. 표2의 데이타 포맷은 표1에 제시된 파형분석결과로 부터 구한 값이다. 이와같이 파형분석과정과 포맷사전이 구축된 후에는 규칙합성단계를 거쳐 합성음을 생성하게 된다. 그림 4와 5는 이와같이 합성된 파형을 나타낸 것이며, 그림 6과 7은 합성음의 포르만트 성분을 추정비교한 것이다.

Ⅳ. 합성음 평가 [1),2),3),4)]

합성음에 대한 평가방법으로는 합성음에 대하여 사전 지식이 없는 남자 1명과 여자 2명을 선정하여 평가표를 작성하여 MOS(Mean Opinion Score)방법으로 구하였다. 평가시 스피치워크스테이션 ver.2.1로 합성한 합성음을 소형 녹음기(AIWA:TP-26)로 녹음한 60개 음절(표 4)을 V,CV,VC,CVC형 순으로 1개 음절씩 번갈아 20초내에 3회 연속 반복음을 청취하여 작성토록 하였다.

표 4. 음절 유형별 합성음 MOS 평가
table 4. MOS test of synthesized speech for the korean syllable tpes

음절 유형	실험대상음절(음절유형별당 15음절 총 60음절)															음질평가			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	이해도	명료도	잡음감	자연성
V형	이	유	우	여	오	어	워	아	의	외	야	요	예	으	에	4.60	4.30	4.20	4.33
CV형	다	리	가	사	지	대	자	가	하	수	무	부	고	기	화	3.46	3.40	3.33	3.42
VC형	일	원	입	연	안	유	양	용	영	엄	열	입	약	역	욱	4.26	3.70	3.65	3.46
CVC형	성	동	상	장	전	불	공	경	분	신	산	신	삶	성	방	3.56	3.30	3.26	3.41

표 3. 데이타 포멧 사전에 등록된 음절표
table 3. syllable table listed in the data format

음절 유형	데이타 포멧 작성 음절(빈도수 순위 확산여 발음대상음)															음절 갯수
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
V형	이	유	우	여	오	어	위	아	의	외	야	요	예	으	어	19 음절
	애	위	와	에												
CV형	다	리	가	사	지	내	자	가	하	수	무	부	고	기	화	80 음절
	따	구	도	주	거	소	차	서	비	조	미	마	세	보	게	
	나	때	내	그	호	휘	개	저	회	교	모	헤	타	바	노	
	끼	재	차	니	세	까	파	짜	카	메	체	배	너	라	싸	
	치	추	무	과	초	트	씨	끼	드	후	포	닷	뷰	피	새	
VC형	일	원	입	연	안	유	안	용	영	업	열	임	약	역	욱	30 음절
	입	운	원	은	불	엄	암	왕	인	원	갑	온	번	옥	입	
CVC형	정	동	상	장	전	불	공	경	분	선	신	신	밝	선	방	100 음절
	관	학	생	중	명	간	단	적	원	감	십	식	한	분	국	
	행	강	반	진	생	잘	확	변	민	중	봄	담	쉬	신	난	
	발	백	천	군	반	쉽	형	편	청	쁨	농	늘	각	말	철	
	낙	복	년	병	결	남	망	삼	목	창	항	식	변	술	충	
	본	건	평	합	숙	곽	쌍	쌍	왕	환	현	밤	숲	삼	책	
	받	합	납	금	담	납	작	넉	색	씽						

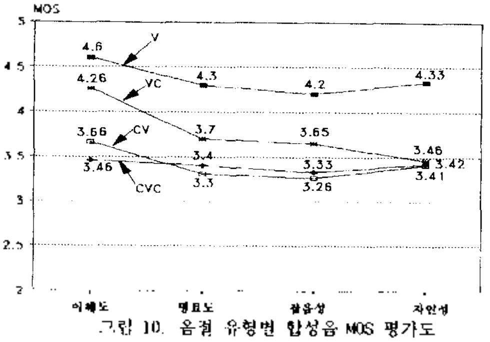

그림 10. 음절 유형별 합성음 MOS 평가도
fig 10. MOS test plot of synthesized speech for the syllable types

표 5. 평가 항목별 MOS 산출표
table 5. MOS evauation table according to the items

	이해도				명료도			잡음감			자연성		
	인지갯수	총점	평균	오인식음	갯수	총점	평균	갯수	총점	평균	갯수	총점	평균
수(5)	26	[illegible]	236/60 =3.96	6개 1. 사(자), 2. 상(장), 3. 산(잔), 4. 신(진), 5. 악(야), 6. 화(바)	15	75	3.60	5	25	3.61	11	55	3.65
우(4)	19	76			25	100		45	180		35	140	
미(3)	9	27			13	39		4	12		8	24	
양(2)					1	2							
총계	54	236		10%	54	216		54	217		54	219	

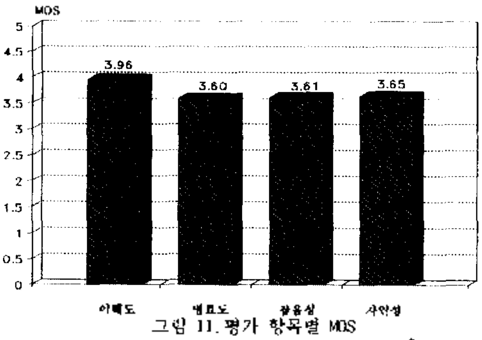

그림 11. 평가 항목별 MOS
fig 11. MOS for the items of evaluation

평가항목은 이해도, 명료도, 잡음감, 자연성등 4가지 항목으로써 5 개의 등급으로 분류하여 등급별로 가산점을 부여하여 평균을 취하였다. 평가시 이해도란은 2개의 항목으로 분류하여 첫번째 항목으로 청취음을 기입하게 하여 이질음화되는 음절을 분석하였으며, 평가시 잘못 청취한 6개의 음절(표 5)은 4가지 평가항목에서 가산점을 영으로 부여하여 평가하였다. 평가등급 항목에서 이해도란은 "5)이해하기 아주 수월하다. 4)이해하기 쉬운편이다. 3)보통이다. 2)이해하기 어렵다. 1)이해하기 아주 어렵다."로, 명료도는 "5)음이 아주 명확하다. 4)원음에 비하여 명확성이 조금 떨어진다.
3) 보통이다. 2) 나쁘다. 1)아주 나쁘다. "로, 잡음감은 "5)잡음이 전혀 없다. 4)좋은 편이다. 3) 보통이다. 2)나쁘다. 1)아주 나쁘다."로, 자연성은 "5)아주 자연스럽다. 4)자연스럽다. 3) 보통이다. 2)어색하다. 1)아주 어색하다."로 평가하였다. 그림 10과 그림 11은 표 4와 표 5를 그림으로 표시한 것이며 음절 유형별로는 V형이 4가지 유형중에서 가장 좋은 결과를 얻었으며, CV형과 CVC형인 경우에는 잡음감과 명료성 항목에서 3.3내지 3.2정도로 평가 항목중 가장 낮은 평가를 받았다. 이질음화되는 합성음도 6개로써 전체 음절 60개의 10%가 이질음화 현상을 보였으나, 대부분 "ㅅ"음을 "ㅈ"음으로 오인식(6개중 4개: 사,상,산,신)하였다(표5). 이질음화가 이루어지는 음절을 유형별로 살펴보면 CV형이 2음절("사"를 "자"로,"화"를 "바"로), CVC형이 3음절("상"을 "장"으로, "산"을 "잔"으로, "신"을 "진"으로), VC형이 1음절("약"를 "야"로) 존재하였다. 그림 11은 평가실험 음절 60개 전체에 대한 평균 음질 이해도는 3.96으로 비교적 합성음을 이해하는 데에는 어려움이 없으나, 명료도, 잡음감, 자연성등은 3.6 정도의 점수로 보통 수준의 결과로써 명료성, 잡음감, 자연성등은 좋은 편에는 미달하는 수준으로 평가되었다.

V. 결론

합성시 25내지 50msec정도의 단구간별로 1 피치주기의 파형을 추출하여 분석된 단음절 파형으로 부터 추출한 진폭 패턴 및 피치 패턴 정보를 이용하여 단음절의 규칙합성음을 생성[5]시켰을 경우에는 명료성 및 음질이 저하 되었다. 본 연구에서는 이러한 결과가 개선되어 합성음을 청취하는 데에는 일반적으로 어려움 없이 이해할 수 있는 수준으로 발전했으며, 잡음감과 명료성도 향상된 것으로 판단되었다. 또한 장단이라든지 강약과 같은 운율변화는 어려움 없이 수행되었으나, 억양을 제어할 수 있는 피치주기의 변화는 이루지 못하였다. 앞으로는 이에 관한 연구가 이루어져야 할 것이다. 또한, 본 논문의 본래 취지인 문어변환 시스템에 적용되기 위하여는 자연스런 무제한 단어합성에 관한 연구가 이루어지야 할 것이다. 이를 위하여는 음절과 음절간의 천이구간에서의 파형처리 및 파형간의 음성학적 고찰과 언어학적 고찰이 병행되어야 음절간에 잘 조화된 자연스런 단어합성음을 생성시킬 수 있을 것이다. 2 음절어 이상의 단어에서 장음과 단음간의 지속시간 비율이라든가 강약정도등과 같은 음성학적 파형특성을 언어학적으로 연계시켜 규칙화시키든지 규칙합성에 필요한 한국어발음사전에 대한 데이타 베이스 작업이 요구된다.

<참고문헌>

1. Nobuhiko Kitawaki, Hiromi Nagabuchi, "Quality Assessment of Speech Coding and Speech Synthesis System," IEEE Comm., 1988. Vol. 26, No.10
2. Toshiro Watanabe, "規則合成音の自然性評價法の檢討",電子情報通信學論文誌,A Vol.J74-A No.4, 1991
3. 조철우, 김경태, 이용주, "무의미단어에 의한 규칙합성음의 평가 및 진단법에 관하여," 음성통신 및 신호처리 워크샵 논문집, 1993.8
4. 김정환, 강성훈, "음성품질 주관법의 표준화에 관한 고찰," 전자통신 동향분석, 1990.7
5. 강찬희, 진용옥, "한국어 문어변환 시스템내에서의 음성합성기 개발," 한국음향학회논문지, Vol.12, No.2, 1993.2