SPEECH SYNTHESIS IN THE TIME DOMAIN BY PITCH CONTROL USING LAGRANGE INTERPOLATION(TD-PCULI)

Chan Hee Kang*°, Yong Jo Shin*, Yun Seok Kim*,
Dae Soo Kang**, Jong Heon Lee**, Ki Hyung Kwon***,
Jeong Keun An***, Sung Tae Sea***, and Yong Ohk Chin***

*, Dept. of Electronics in Sangji junior college
**, Telesys ERC Telematic System Engineering Research Center
***, Dept of Electronic Engineering in Kyunghee Univ.

<ABSTARCT>

In this paper a new speech synthesis method in the time domain using mono-syllables is proposed. It is to overcome the degradation of the synthetic speech quality by the synthesis method in the frequency domain and to develop an algorithm in the time domain for the prosodic control. In particular when we use a method in a time domain with mono-syllable as a synthesis unit it will be the main issues which are to control the pitch period and to smooth the energy pattern. As a solution to the pitch control, a method using Lagrange interpolation is suggested. As a solution to the other problem, an algorithm which can control the amplitude envelop shape of mono-syllable is proposed. As the results of experiments it was possible to synthesize unlimited Korean speeches including the prosody control. Accoding to the MOS evaluation[1)-5)] the quality and the naturality in them was improved to be a good level.

I .Introduction

In this paper we present a new method to control the prosodic factors such as a stress, a duration and a intonation to be maintained original sounds. This is almost similar with TD-PSOLA method in the facts that it uses the pitch-synchronization technique, it is non-parametric method, its hot issues are to control the prosodic factors without the degradation of the voice qualities, the clearness of synthetic speeches is superior to other methods and so on. TD-PSOLA method overlays and adds hanning window function with 2T pitch periods to about 10msec ST(Short-Time) intervals to control

the intontion. And it controls the duration by selecting the marked staionary part speech data in stored CDUs. But in this paper we propose another method which can control the intonation by alternating each pitch period frames with new pitch period frames using a interpolation technique without overlaying intervals and the duration by the decimation or the repetition of pitch period frames.

Ⅱ. The algorithm for the synthesis -by-rule[1)]

Fig. 1 is represented by a overall block diagram for the synthesis-by-rule. The process is illustrated as follows. First, we extract parameters for the control of prosodic factors from stored speech data through the

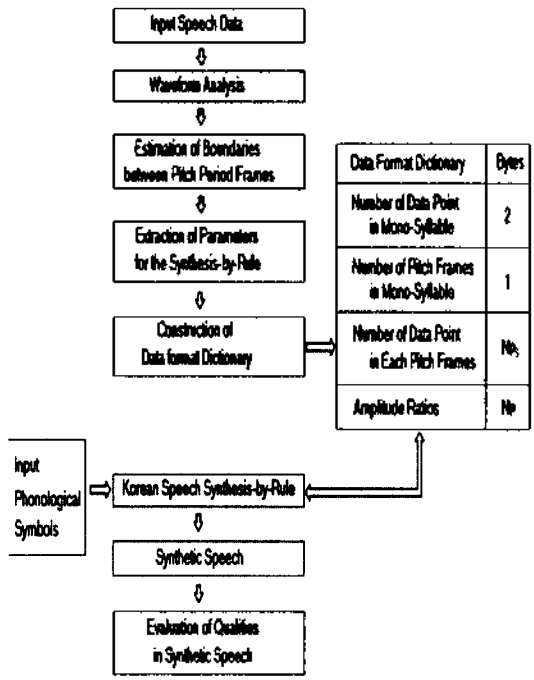


fig.1The overall block diagram of synthesis-by-rule.

table 1. The waveform analysis table

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

No.	DATA POINT (PTS)	MAX (mV)	INTERVAL (PTS)	DATA POINT (PTS)	MIN (mV)	INTERVAL (PTS)	No. CROSSING	ZERO	MAX RATE	MIN RATE
1	496	55	95	457	-23	70	1		0.30	-0.12
2	554	132	70	543	-116	85	2		0.71	-0.63
3	635	173	83	626	-187	84	4		0.93	-1.00
4	721	151	88	711	-165	86	6		0.81	-0.88
5	810	97	90	796	-139	88	4		0.52	-0.74
6	901	96	90	887	-156	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	-131	89	2		0.46	-0.71
11	1348	82	90	1330	-127	90	2		0.44	-0.68
12	1437	82	89	1419	-111	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	-0.28
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	2411	-10	71	2		0.08	-0.05

waveform analysis process. Table 1 is given the parameters extracted through the process. These prameters are registered into the data format dictionary[1)-3)]. Then synthetic speeches are generated by using them in the dictionary according to the Korean phonological rules[1)].

table 2. An example of data format dictionary

No.	FORMAT NAME	DS	DC	NTOTAL	NP	NA1	1 PERIOD	AMP. RATIO
1	I.FDA	0	0	2099	22	912	128 99 ...	87 76 .2 .5 ... .2 .1
2	DA.FDA	1	186	1674	16	2304	114 89 ...	93 73 .6 .5 ... .1 .1
3	ID.FDA	0	0	2488	32	2348	95 79 ...	68 52 .3 .4 ... .1 .0
4	JEONG.FDA	1	434	2392	22	2716	106 80 ...	89 69 .3 .8 ... .1 .1
5	LI.FDA	0	0	2000	22	2068	103 88 ...	92 82 .3 .2 ... .1 .1
6	MYEON.FDA	0	0	2736	33	3320	111 86 ...	72 66 .2 .4 ... .1 .1
7	BONG.FDA	1	287	2144	21	2672	86 78 ...	85 74 .3 .8 ... .1 .1
8	IL.FDA	0	0	2390	30	2464	92 77 ...	79 81 .2 .4 ... .1 .1
9	SANG.FDA	1	508	3940	27	2956	114 79 ...	89 75 .3 .8 ... .1 .1
10	GI.FDA	1	497	1884	15	1884	99 87 ...	92 77 .6 .7 ... .1 .1
11	SA.FDA	1	488	2350	19	3048	109 84 ...	89 80 .6 .8 ... .1 .1
12	JI.FDA	1	530	2142	18	1836	92 85 ...	90 85 .3 .5 ... .1 .1
13	MU.FDA	0	0	2298	26	1528	106 80 ...	74 63 .2 .6 ... .2 .1
14	JEON.FDA	1	475	2624	24	3052	113 83 ...	84 67 .5 .9 ... .1 .1
15	O.FDA	0	0	1642	17	2176	125 91 ...	99 70 .1 .3 ... .1 .1
16	GUK.FDA	0	0	2742	34	1872	109 82 ...	78 76 .1 .1 ... .1 .0
17	MOL.FDA	0	0	1920	22	2644	73 80 ...	85 70 .1 .3 ... .2 .1
18	EO.FDA	0	0	1654	18	2152	104 88 ...	85 74 .2 .4 ... .1 .1
19	GONG.FDA	1	490	2838	23	2448	78 85 ...	93 71 .1 .6 ... .1 .1
20	JA.FDA	1	424	2112	19	2328	98 79 ...	84 57 .2 .8 ... .1 .0
21	GA.FDA	1	310	2024	19	3084	103 90 ...	82 57 .4 .6 ... .1 .0
22	MOM.FDA	0	0	2540	28	3300	137 92 ...	93 51 .1 .3 ... .1 .0
23	A.FDA	0	0	1592	18	2438	99 57 ...	91 72 .1 .3 ... .1 .1
24	BA.FDA	1	535	2086	18	3468	109 77 ...	81 48 .2 .5 ... .1 .0
25	SEON.FDA	1	694	2592	22	3612	110 81 ...	54 17 .5 .9 ... .0 .0
26	EUI.FDA	0	0	2244	24	1512	108 88 ...	87 70 .2 .4 ... .2 .1
27	SU.FDA	1	1512	2932	16	3240	91 79 ...	90 115 .4 .6 ... .2 .1
28	EOP.FDA	0	0	1090	13	2696	107 84 ...	89 55 .1 .2 ... .2 .0
29	EO.FDA	1	329	1862	17	3204	129 84 ...	90 42 .5 .8 ... .1 .0
30	IN.FDA	0	0	2478	31	2048	102 80 ...	75 87 .2 .4 ... .2 .1

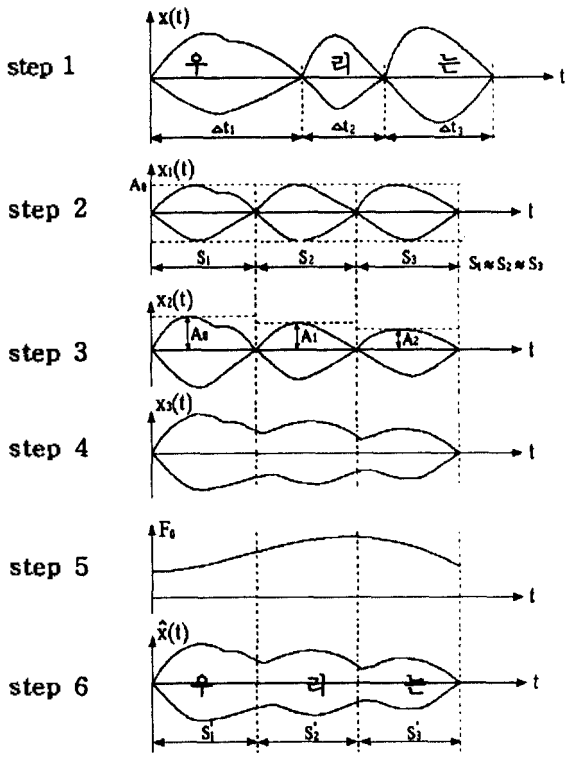


fig. 2 The process plot for the synthesis by rule.

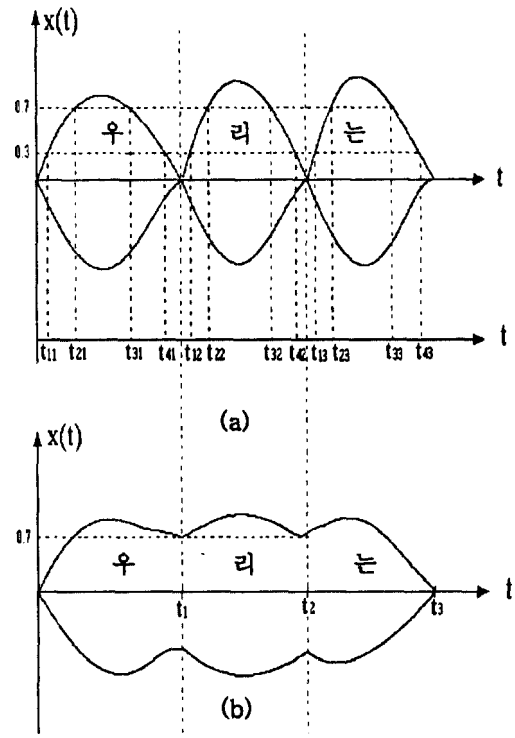


fig. 3 The process plot to control the energy pattern.

Table 2 represents an example of data format dictionary. Amplitudes, durations and intonations of each mono-syllables stored into the memory are different from each other. So that, we have to control them according to the Korean phonological rules to synthesize the speeches. Fig. 2 represents the step-by-step processes for this. Step 1 represents the amplitude normalization process of each mono-syllables. After the normalization it begins with the next process to control the duration by doing the decimation or repetition of 1 pitch period frames as in the step 2. And then we control the stress to be a constant amplitude(i.e., A_0, A_1, and A_2 in fig. 2<step 3>) by weighting to them. Step 4 is illustrated in fig. 3. This process is to control the energy pattern in continuous synthetic speeches. First, shape patterns of syllables to modify have to be decided by the coupling rules[1] of between syllables. Then we have to change a shape of waveform according to the rules. In fig.3 marked points such as t_{1i}, t_{2i}, t_{3i}, and t_{4i} mean parts of each syllables to truncate it off or to change it into another shape pattern. Then shpae pattern of continuous speeches are produced as in fig. 2<step 4> and fig. 3(b). Step 5 in fig. 2 represents the process to add pitch period pattern into continuous speeches. And fig. 4 means the step-by-step process to modify the pitch period pattern of original speech into another pattern which can be

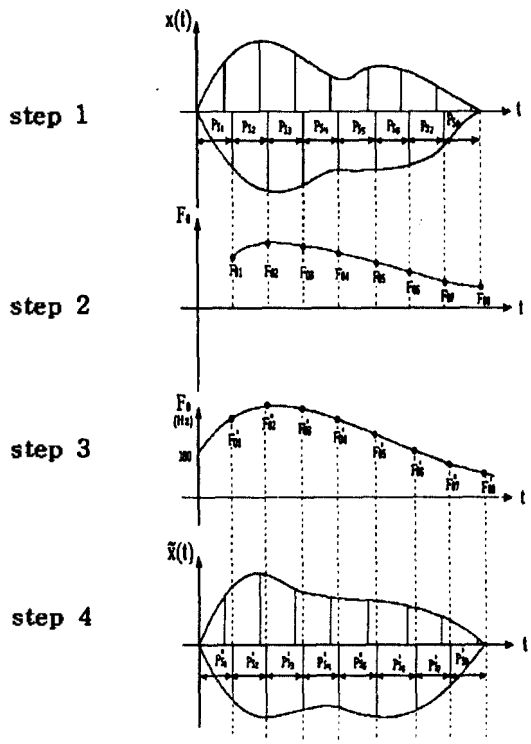


fig.4 The process plot for the modification of pitch period pattern in a syllable unit.

decided according to input text structures[5]. Step 1 means the process for the extraction of pitch period frames from the stored speech data through the waveform analysis. Step 2 is represented by the fundamental frequency of mono-syllable shown in fig4<step 1> which is to be changed into new one such as shown in fig.4<step 4>. We can control the intonation by the method that 5 order Lagrange interpolations are performed on each pitch period frames. The results of this are shown in fig. 5. In the pitch control only the region which is to be observated the pseudo-periodic characteristics in the stored original speech is interpolated with

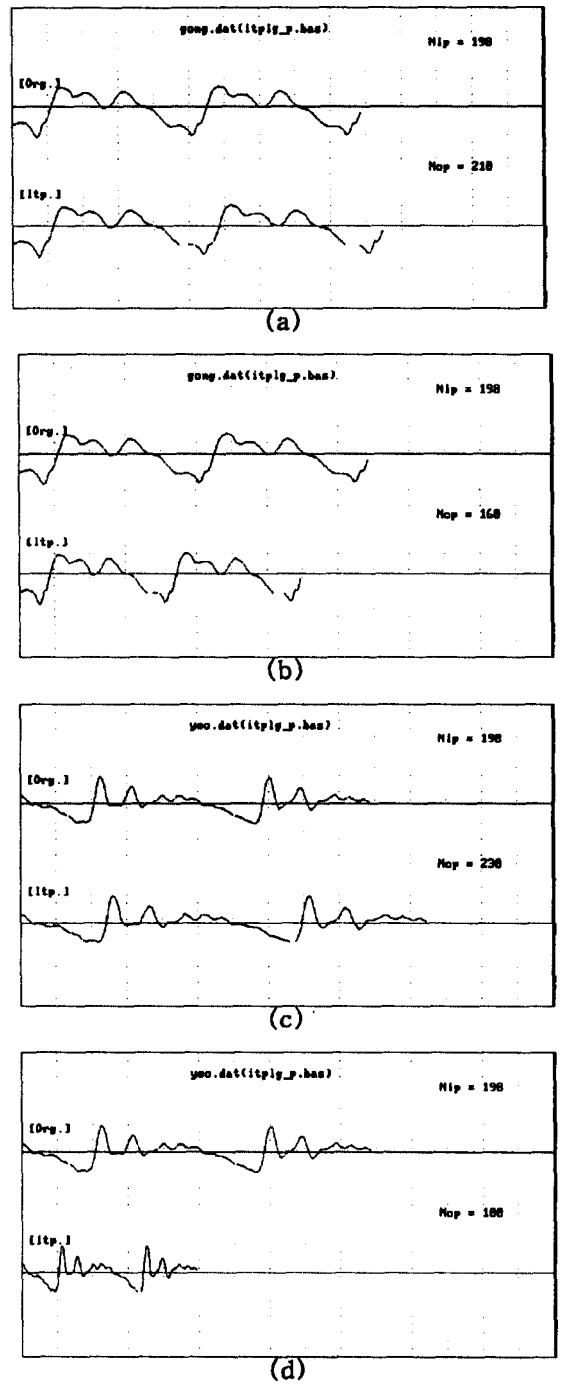
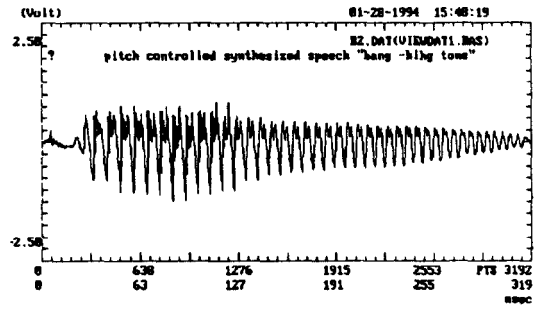


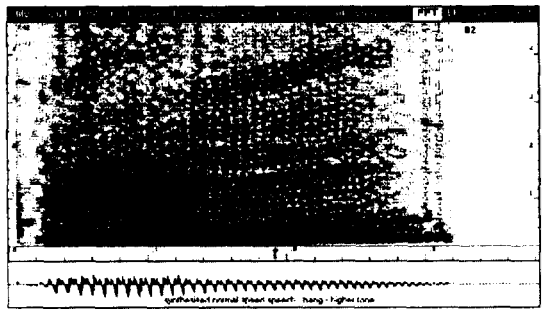
fig. 5(a),(b),(c),(d) Examples of shrinking and expansion of pitch periods

a Lagrange interpolation. Fig.5(a) and (b) are the interpolated examples of the Korean CVC-type syllale "gong" and fig. 5(c) and (d) are the examples of double vowel "yeo".

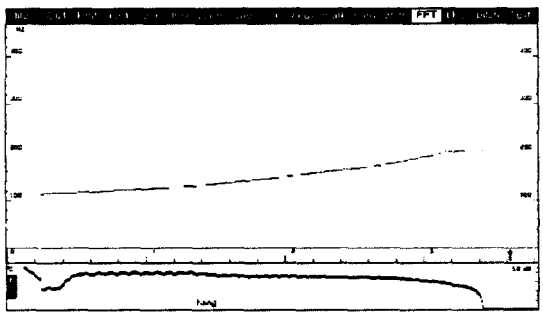
Ⅲ. Examples of the pitch control using Lagrange Interpolation



(a)

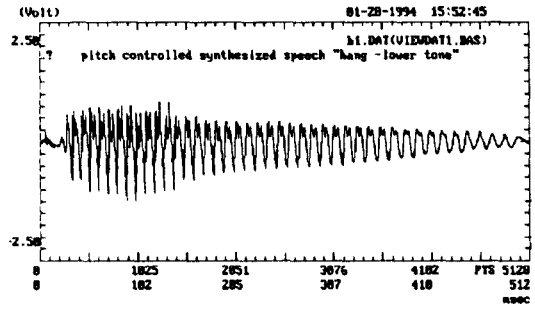


(b)

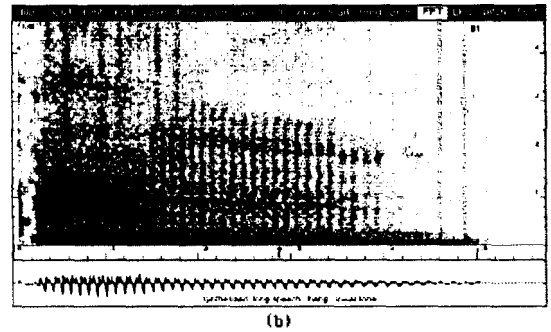


(c)

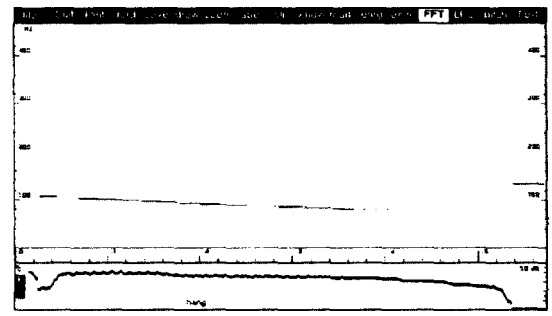
fig. 6(a),(b),(c) An example of mono-syllable changed into rising intonation("bang" : 0.319 sec)



(a)

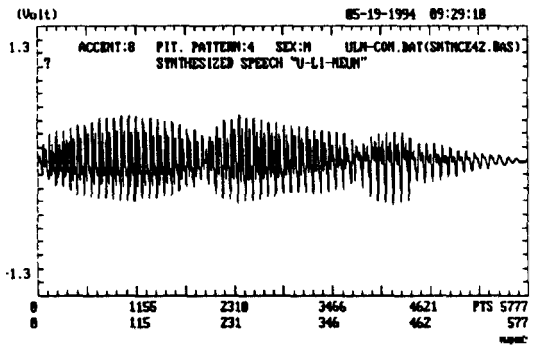


(b)



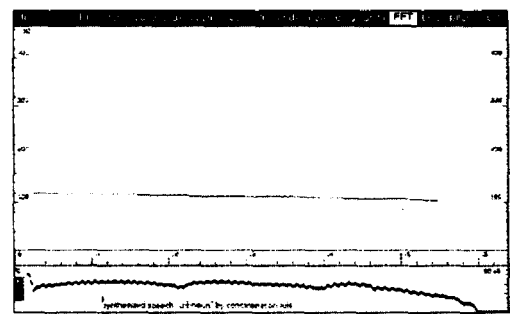
(c)

fig. 7(a),(b),(c) An example of mono-syllable changed into falling intonation("bang" : 0.513 sec)



(a)

fig. 8(a) An example of controled namplitude envelope



(b)

fig. 8(b) An example of controled pitch patter

Ⅵ. Conclusion

In this paper a new speech synthesis method in time domain using mono-syllables is proposed. It is to overcome the degradation of the quality in synthetic speeches by the synthesis method in the frequency domain and to improve the naturality by the method in the time domain which is difficult to control the prosodic factors. In general it will be the main issues which are to control the prosodic factors and the recovery to the degradation of the naturality in the synthetic speeches due to the articulation in the case of synthesizing the speech by the time domain using the mono-syllables as the synthesizing unit. As a solution to the pitch control, a method that it is to synthesize the speeches using the parameters for the prosodic controls extracted from them after the sharing of the unit pitch frames by searching the maximum point in a pitch period frame(i.e., 1T pitch period) generated whenever the vocal cords are vibrated is suggested. As a solution to the other problem, an algorithm which can control the amplitude envelop shape of mono-syllable in coupling regions between the mono-syllables is proposed. It seems to be far from reaching at the perfect method to process the Korean allophones which changes with a variety, but according to the MOS evaluation the naturality in continuous synthetic speech is improved.

As the results of experiments to the synthetic speeches[1)-5)] it is a possible method to synthesize unlimited Korean speeches and seems to be improved the quality and naturality in them. We will propose it to be called as a TD-PCULI(Time Domain-Pitch Control using Lagrange Interpolation) method.

<References>

1. C. H. Kang, "A Study on the Korean Speech Syntehsis-by-Rule using Unit Pitch Frame Informations," Kyunghee Univ. Doctorate Thesis, 1994. 8

2. C. H. Kang, Y. O. Chin, "Development of Speech Synthesizer in Korean TTS System," The Journal of the Acoustical Society of Korea, Vol. 12, No. 2, 1993.2

3. C. H. Kang, Y. O. Chin, "Speech Synthesis for the Korean large Vocabulary Through the Waveform Analysis in Time Domains and Evauation of Synthesized Speech Quality," The Journal of the Acoustical Society of Korea, Vol. 13, No. 1, 1994.

4. C. H. Kang, J. H. Lee, J. K. An, K. H. Kwon, S. T. Sea., Y. O. Chin, "The Evaluation of Speech Quality Synthesized by Rule to Korean Syllable Types," 1993 Conference of the Acoustical Society of Korea, Vol.12,

No.1, 1993. 12.
5, C. H. Kang, Y. O. Chin, Ohers, "Evaluation of the Synthetic Speech Quality by the TD-PCULI method," WESTPRAC V proc., 1994. 8
6. Jonathan Allen, M. Sharon Hunnicutt and Dennis Klatt, From Text to Speech : The MITalk system, Cambridge Univ. Press, 1987
7. Shuzo Saito, Fundamentals of Speech Signal Processing, Academic Press, 1981
8. G. Rigoll, "The DECtalk system for German : A study of the modification of a text-to-speech converter for a foreign language," IEEE Proc. ICASSP '87, 1987
9. Nobuhiko Kitawaki, Hiromi Nagabuchi, "Quality Assessment of Speech Coding and Speech Synthesis System," IEEE Comm., 1988. Vol26. No.10
10. Toshiro Watanabe, "規則合成音の自然性評價法の檢討", 電子情報通信學 會論文誌, A Vol. J74-A No.4, 1991
11. J. H. Kim, S. H. Kang, "A Study on the Standardization of Subjective Assessment of Speech Quality," Electronic Communication Movement Analysis, 1990.7
12. C. H. Kang, B. Y. Kim, Y. O. Chin, "Synthesis of the Korean Syllables using the Unit Pitch Frame Informations," 1988 Summer Conference of the Korean Institute of Communication Sciences, 1988. 8.