

HDB-3 데이터전송처리 효율을 고려한 7비트 한글 자모 부호체계

7bits Hangul Jamo Coding system to consider of HDB-3 Data transmission processing efficiency

홍완표*

Wan-Pyo Hong*

요 약

본 논문은 3비트 열, 4비트 행 7비트 1바이트 부호체계를 갖는 한글정음부호체계에 대하여 데이터 전송효율 측면에서 분석하였다. 그 결과 한글정음부호체계는 원천부호화 규칙에 위배되는 부호의 총 사용빈도가 19.78%에 달하여 이 부호체계를 사용할 경우 데이터의 전송효율을 상당히 저해할 것으로 나타났다. 따라서 본 논문에서는 그에 대한 대책으로 원천부호화 규칙에 부합되는 새로운 한글정음부호체계를 제시하였다.

Abstract

This paper proposed a new Hangul Jeongum Jamo coding system. To do it, the current Hangul Jeongum coding system was analyzed with the source coding rule of character. As a result of the study, the current Hangul Jeongum coding system should be changed because there are many violated codes in the coding system. When the proposed coding system for Hangul Jeongum coding system is applied, the datacommunication efficiency could be increased to 19.78%.

Key words : Jeongum 3bul 3bit source code, Hangul 3bul coding, Hangul 3bul frequent

I. 서 론

정보기기의 키보드에서 입력되는 문자나 숫자는 2진 비트로 부호화되어 처리된다.

이러한 문자나 숫자는 7비트 또는 8비트 단위로 부호화된다. ASCII부호는 대표적인 7비트 부호체계이다. EBCDIC부호는 대표적인 8비트 부호체계이다. 한글은 참고문헌 [1]의 부속서 4에 7비트 한글날자 부호계에 대한 권장 부호체계를 제시하고 있다. 이 부호체

계는 정보 교환용 부호계를 내부적으로 사용하지 못하는 시스템과 관련 장비에서 사용하는 한글 부호계의 표현 형식에 대한 것이다. 또한 이 부호체계는 정보 교환용으로 사용하지 않는 것을 원칙으로 한다. KS C5601-1987(Multi-byte Graphic Character set for Information Exchange)와 RFC 1557(Korean Character Encoding for Internet Message)은 정보교환용 7비트 부호체계이다. 전자는 7비트열과 7비트 행으로 구성되는 14비트 2바이트 부호체계이다. 후자는 전자를 인

* 한세대학교 정보통신공학과(Dept. of Information & Telecommunication Hansei University)

· 제1저자 (First Author) : 홍완표 (Wan-Pyo Hong, Tel : +82-11-380-0685, email : seoulhong@paran.com)

· 투고일자 : 2012년 10월 4일

· 심사(수정)일자 : 2012년 10월 6일 (수정일자 : 2012년 12월 21일)

· 게재일자 : 2012년 12월 30일

용한 것이다. 참고문헌 [1]의 부속서 3 표2에는 한글 낱자에 대한 부호값을 제시하고 있다. 이 표에서 제시하고 있는 한글낱자의 부호 값은 5비트 부호체계이다. 따라서 현재 데이터 교환용으로 제시된 7비트 부호체계의 한글 낱자는 KSC 5601-1987이다. 미래형한글문자판 포럼 다 언어 치환 자판배열 표준화위원회에서는 다 언어를 국제음성기호로 치환한 후에 그것을 정음부호로 변환하여 송신하는 것을 제안하였다. 이 정음부호체계는 초성, 중성 및 종성으로 구분되는 한글 낱자 부호체계로서 7비트 부호체계이다.

이러한 부호체계들은 대부분 한글의 자음과 모음의 순으로 또는 자음과 모음을 조합한 글자의 순서로 부호화되었다. 본 논문에서는 다 언어 치환 자판배열 위원회에서 제시한 7비트 한글정음낱자 부호체계를 중심으로 하여 데이터 전송효율을 고려한 한글정음 부호체계를 제시하였다. 7비트 부호체계의 데이터 전송효율은 참고문헌[2]에서 제시한 3비트 열, 4비트 행 원천부호화 규칙을 적용하였다.

II. 한글낱자 7비트 부호체계 현황

2-1 KS X 1001 부속서 4

이 부호체계는 정보 교환용 부호계를 내부적으로 사용하지 못하는 정보기기에서 사용하는 한글 부호계의 표현 형식을 나타내고 있다.

또한 이 부호체계는 정보 교환용에 적용하지 않는 것을 원칙으로 한다. 그러므로 부호체계에 대한 데이터전송효율을 고려하는 본 논문에서는 이 부호체계에서 권장하는 현황만 소개하였다. 이 부호체계는 7비트 한글 낱자 1자를 표현하고 있다. 이 권장 안에서 제시한 한글 낱자 부호 체계는 표 1과 같다. 이 표1에서 보는 바와 같이 이 부호체계는 한글 낱자를 초성, 중성 및 종성으로 구분하지 않고 자음과 모음으로 구분하고 있다. 총 한글낱자는 51개이며 채움부호까지 총52개의 부호로 구성되어 있다. 이 부호체계에서 총 가용부호는 64개이다.

표 1. 7비트 한글 낱자 부호계
Table 1. Hangul Jamo 7bit coding system

구 분				b7	0	0	0	0	1	1	1	1
				b6	0	0	1	1	0	0	1	1
				b5	0	1	0	1	0	1	0	1
b4	b3	b2	b1	0	1	2	3	4	5	6	7	
0	0	0	0	0	기능문자			채움	ㄱ			
0	0	0	1	1				ㄴ	ㅋ			
0	0	1	0	2				ㄷ	ㆁ	ㅌ	ㄷ	ㄱ
0	0	1	1	3				ㄹ	ㅃ	ㅆ	ㅅ	ㅅ
0	1	0	0	4				ㄴ	ㅅ	ㅈ	ㅈ	ㄱ
0	1	0	1	5				ㄴ	ㅅ	ㅈ	ㅈ	ㅅ
0	1	1	0	6				ㄴ	ㅅ	ㅈ	ㅈ	ㅅ
0	1	1	1	7				ㄴ	ㅅ	ㅈ	ㅈ	ㅅ
1	0	0	0	8				ㄴ	ㅅ			
1	0	0	1	9				ㄴ	ㅅ			
1	0	1	0	10				ㄴ	ㅅ	ㅈ	ㅈ	ㅅ
1	0	1	1	11				ㄴ	ㅅ	ㅈ	ㅈ	ㅅ
1	1	0	0	12				ㄴ	ㅅ	ㅈ	ㅈ	ㅅ
1	1	0	1	13				ㄴ	ㅅ	ㅈ	ㅈ	ㅅ
1	1	1	0	14				ㄴ	ㅅ	ㅈ	ㅈ	ㅅ
1	1	1	1	15			ㄴ	ㅅ	ㅈ	ㅈ	ㅅ	

* 4/0 채움 문자의 용법: 한 글자마디는 세 바이트로 나타내며, 첫소리, 가운뎃소리, 끝소리 글자가 없을 때, 그 자리에 각각 채움 문자를 넣는다.

표 2. 한글정음 부호체계
Table 2. Hangeul Jeongum Coding system

열 \ 행	4	5	6	7
0	(초성)	ㅇ	(중성)	ㅇ
1	ㄱ	ㅇ	ㄱ	ㅇ
2	ㄴ	(중성)	ㄴ	(중성)
3	ㄷ	ㅏ	ㄷ	·
4	ㄹ	ㅑ	ㄹ	—
5	ㅍ	ㅓ	ㅍ	1
6	ㅂ	ㅕ	ㅂ	
7	ㅅ	ㅗ	ㅅ	
8	ㅇ	ㅛ	ㅇ	
9	ㅈ	ㅜ	ㅈ	
10	ㅊ	ㅠ	ㅊ	
11	ㅋ		ㅋ	
12	ㅌ		ㅌ	
13	ㅍ		ㅍ	
14	ㅎ		ㅎ	
15	△		△	

표 2는 미래한글자판포럼 다언어치환자판위원회에서 제시한 한글정음부호체계이다. 이 부호체계에서 보듯이 한글낱자를 초성, 중성 및 종성으로 구분하고 있다. 이 부호체계에서는 표 1의 부호체계와 달리 한글 옛자 4자를 포함하고 있다. 이 부호체계는 총 가용 부호 64개 중에서 51개를 사용하고 있다. 이 부호체계는 표1과는 달리 채움문자를 적용하고 있지 않다. 이 부호체계는 부호표안에서의 낱자의 배열을 표1과 같이 자음과 모음의 순으로 하고 있다.

III. 데이터 전송효율과 정음부호체계

3-1 데이터 전송과 부호체계

정보기에 입력되는 문자는 일정한 형식을 가진 2진 비트로 부호화된다. 이러한 부호화를 원천부호화(source coding)라고 한다. ISO는 데이터통신의 계층을 7계층으로 구분한다. 원천부호화는 이 계층의 6계층인 표현계층에서 행하여진다. 이와 같이 정보기 내부에서 부호화된 문자의 원천부호는 물리계층을 거쳐

전송로를 통해 원거리에 보내진다. 원천부호는 물리계층에서 전송로에 송출되기 전에 전송로에 적합한 신호로 변환된다. 물리계층에서의 이러한 신호의 변환을 회선부호화라고 한다. 문자의 원천부호는 전송에 적합하지 않은 일정한 개수 이상의 2진 비트 “0”을 가질 수 있다.

이것은 원천부호가 전송신호로 변환되어 전송로에서 전달될 때 전체 신호품질에 악영향을 주게 된다. 또한 이것은 수신기에서 전송신호를 2진 비트로 복호화하는데 지장을 줄 수 있다. 즉 수신기에서 원하는 정보를 복원하지 못하게 하는 원인이 된다. 그러므로 물리계층에서는 원천부호에 이러한 일정 개수 이상의 연속되는 “0”의 비트가 있을 때 이것을 다른 부호체계로 바꾸어 회선부호화하게 된다. 이러한 것을 스크램블링이라고 한다. ITU-T와 한국에서는 이러한 스크램블링 기술로 HDB-3방식을 표준으로 채택하고 있다. HDB-3방식은 원천부호에 연속되는 4개의 2진 “0”의 비트가 있을 때 이것을 연속“0”의 비트가 아닌 정해진 패턴으로 변환시켜 준다. 그러므로 원천부호에 연속 4개 이상의 2진 비트 “0”이 많을 경우, 물리계층에서의 데이터의 처리 효율을 떨어뜨리게 된다. 즉 이것은 데이터의 전송효율을 떨어뜨리게 하는 원인이 된다.

3-2 원천부호화 규칙

가. ECMA 원천부호화 규칙

표 3은 ISO의 Standard ECMA-35 그림 9에서 제시하고 있는 원천부호화를 위한 기준이다 [3]. 표 3에서 CL(Control location)은 장치의 제어에 대한 부호영역이다. 그리고 GL(Graphic location)은 문자나 숫자 등에 대한 부호영역이다. 02/00은 “space” 부호, 07/15는 “delete” 부호영역으로 고정된 영역을 가지고 있다. 이 표 3에서 보듯이 이 기준은 문자의 원천부호를 위한 부호영역을 제시하고 있다. 원천부호화를 하는 경우에 이 각각의 영역 내에 해당되는 문자나 숫자 등을 배열하는 방법에 대하여는 그 기준을 제시하고 있지 않다. 위에서 언급한 바와 같이 현재는 이 각각의 영역 내에서 특별한 기준이 없이 영어의 경우에는 알파벳 순서로, 한글의 경우에는 자모순으로 배열하고 있다.

표 3. 문자부호구조와 부호화 기법
Table 3. Character Code Structure and Extension Techniques

	00	01	02	03	04	05	06	07	
00	CL		SP						
01									
02									
03									
04									
05									
06									
07				GL					
08									
09									
A									
B									
C									
D									
E									
F								DL	

나. 문자의 원천부호화 규칙

표 4와 표 5는 3비트 열, 4비트 행 7비트 원천부호에 대한 부호화 규칙이다[10]. 표 4는 각 문자, 기호 또는 숫자의 원천 부호화하는 규칙이다. 표 4에서 상위 비트 열은 표 1의 b5~b7을 구성하는 비트 열이다. 이것이 있는 아라비아숫자는 16진수이다. 표 4에서 하위 비트 열은 표 1의 b1~b4를 구성하는 비트 열이다. 이것에 있는 아라비아숫자는 상위 비트 열에서와 같이 16진수이다. 표 1과 표 2에서 상위 비트 열은 16진수 4, 5, 6과 7만 사용하고 있다. 표 5는 바이트 간 조합될 때를 고려한 것이다. 첫 번째 바이트의 상위비트 4비트를 기준으로 하여 이어지는 두 번째 바이트의 하위비트 3비트에 대한 것이다.

표 4. 문자부호화 규칙; 3비트 열 x 4비트 행
Table 4. Character coding rule; 3bits column 4bits row

상위 비트 열 (2진수, 3비트)	하위 비트 행 (16진수, 4비트)	
	부여 제한	부여 가능
000	0, 1, 2, 3, 4, 5, 6, 7	8, 9, A, B, C, D, E, F
001	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
010	0, 1	2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
011	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
100	0, 1, 2, 3	4, 5, 6, 7, 8, 9, A, B, C, D, E, F
101	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
110	0, 1	2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
111	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F

3-3 한글정음부호체계와 원천부호화 규칙

표 2의 한글정음부호체계의 구조는 표 3의 국제부호체계 구조에 부합된다. 한글정음부호체계를 데이터 전송효율을 고려한 문자 원천부호화 규칙인 표 4와 표 5에 적용할 경우 표 6과 같은 결과를 보여 주었다. 이 결과에 의하면 한글정음부호체계가 표 5와는 부합되었다. 그러나 표 4와는 표 6과 같이 부합되지 않는 부분이 있었다. 표 6에서 보는 바와 같이 초성의 “ㄱ ㄴ ㄷ ㅇ” 종성의 “ㄱ” 초성의 “ㅇ” 부호가 표 4의 원천부호화 규칙에 맞지 않는 것으로 나타났다.

표 5. 바이트 (7비트)간 조합규칙 ; 첫 번째 바이트, b7-b5 기준

Table 5. Coding rule between bytes(7bits); based of b7-b5 of First byte

두 번째 바이트, b4-b1	첫 번째 바이트, b7-b5	부호 부여 제한	부호부여가능
0000	000	0, 2, 4, 6, 8, A, C, E	1, 3, 5, 7, 9, B, D, F
0001	001	0, 4, 8, C	1, 2, 3, 5, 6, 7, 9, A, B, D, E, F
0010	010	0, 8	1, 2, 3, 4, 5, 6, 7, 9, A, B, C, D, E, F
0011	011	0, 8	1, 2, 3, 4, 5, 6, 7, 9, A, B, C, D, E, F
0100	100	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
0101	101	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F
0110	110	0	1, 2, 3, 4, 5, 6, 8, 9, A, B, C, D, E, 7, F
0111	111	0	1, 2, 3, 4, 5, 6, 7, 8, 9, A, C, D, E, B, F

이와 같이 원천부호화 규칙에 위배되는 부호들이 문자 전체에서 어느 정도의 사용빈도를 차지하고 있는지를 분석할 필요가 있다. 그 결과에 따라 현재의 한글정음부호 체계를 그대로 사용할 것인지 수정할 것인지를 판단하여야 한다. 분석결과 해당 문자부호들의 사용빈도가 낮을 경우에는 현재의 한글정음부호 체계를 그대로 사용하여야 할 것이다. 문자의 부호화 방식에서 문자의 사용 빈도율을 사용하고 있는 대표적인 방식은 모르스 부호화방식과 허프만 부호화방식이다[5]. 이 두 개의 부호화 방식을 엔트로피(entropy) 부호화방식이라 한다. 이 방식들은 상대적으로 사용 빈도가 높은 문자에 상대적으로 짧은 부호를 부여하

는 방식으로 평균정보량을 작게 하고 있다. 즉 전송정보량을 감소시킴으로서 전송효율을 높이는 방식이다. 본 논문에서는 상대적으로 사용빈도가 높은 문자의 부호에 일정개수 이상의 연속 “0”의 비트 열이 발생되지 않는 문자부호를 부여하는 방식이다. 즉 이로써 전송되는 문자 부호 중에서 일정 개수 이상의 비트“0”의 발생을 감소시킴으로서 스크램블링으로 인한 전송효율이 감소되는 적을 낮추는 것이다. 표 7은 한글날자의 초성, 중성 및 종성에 대한 사용빈도를 나타내는 것이다.

표 6. 한글정음부호체계와 원천부호화 규칙

Table 6. Hangeul Jeongum code system and Source coding rule

열 \ 행	4	5	6	7
0	(초성)	ㅇ	(종성)	ㅇ
1	ㄱ	ㅇ	ㄱ	ㅇ
2	ㄴ	(중성)	ㄴ	(중성)
3	ㄷ	ㅏ	ㄷ	·
4	ㄹ	ㅑ	ㄹ	ㅡ
5	ㅓ	ㅓ	ㅓ	ㅣ
6	ㅕ	ㅕ	ㅕ	
7	ㅗ	ㅗ	ㅗ	
8	ㅛ	ㅛ	ㅛ	
9	ㅜ	ㅜ	ㅜ	
10	ㅠ	ㅠ	ㅠ	
11	ㅋ		ㅋ	
12	ㅌ		ㅌ	
13	ㅍ		ㅍ	
14	ㅎ		ㅎ	
15	△		△	

이 표 7에서 보는 바와 같이 한글정음 부호체계에서 문자 원천부호화규칙에 위배되는 각 문자부호에 대한 전체 낱자 중 총 사용 빈도율을 보면 다음과 같다. 단, 초성과 종성의 옛글자를 제외하였다.

이 결과에 의하면 한글정음부호체계에서 원천부호화 규칙에 위배되는 문자부호의 총 사용 빈도율은 초성 ㄱ 5.89% ㄴ 2.44% ㄷ 10.35% 총 18.68%, 종성 ㄱ 1.10% 합계 19.78%에 달한다. 그러므로 이 원천부호화 규칙에 위배되는 문자부호들을 사용빈도가 낮은 문자부호와 대체할 경우 회선부호기에서 그에 상당하는 스크램블링 횟수를 줄이게 됨으로 그 만큼 데이터의 전송효율을 높일 수 있게 되는 것이다. 표 7에서 보듯이 초성의 경우 한글정음부호체계에서 사용되는 낱자 중 “ㄱ”이 사용빈도가 가장 낮다. 종성의 경우에는 “ㅍ”이고 종성의 경우에는 초성과 마찬가지로 “ㄱ”

이다.

3-4 한글정음부호체계 개선방안

이상에서 분석한 바와 같이 현재의 한글정음부호체계는 그 구조에 있어서는 표 3의 국제부호화 구조에 부합된다. 그러나 표 4의 원천부호화 규칙에는 일부 위배되는 부호가 있는 것으로 나타났다. 따라서 그 위배된 부호들을 원천부호화 규칙에 부합되도록 부호화할 경우에 최대 19.78%의 개선효과를 얻을 수 있다. 그러나 어떠한 방법으로 부호체계를 개선하느냐에 따라 그 개선효과는 달리 나타날 수 있다. 본 연구에서는 표 8, 표 9 및 표 10에서 보여 주는 것과 같이 세가지의 개선방안을 제시하였다. 첫 번째 방안으로는 기

표 7. 한글낱자 사용빈도수, 빈도율 및 총 빈도율

Table 7. Hangeul Jamo frequent number, frequent rate and total frequent rate

초성	hexa	빈도수	빈도율 (%)	총빈도율 (%)	중성	hexa	빈도수	빈도율 (%)	총빈도율 (%)	종성	hexa	빈도수	빈도율 (%)	총빈도율 (%)
ㄱ	1100	308480	23.45	10.27	ㅏ	1161	535428	41.54	17.86	ㄴ	11AB	88733	22.32	2.98
ㅇ	110B	232752	17.69	7.75	ㅣ	1175	187794	14.57	6.26	ㄷ	11AF	72026	18.12	2.49
ㄱ	1100	175213	13.32	5.83	ㅑ	1165	132439	10.27	4.42	ㅇ	11BC	66552	16.74	2.30
ㅏ	1109	101877	7.74	3.39	ㅓ	1169	122451	9.50	4.08	ㅕ	11BB	36755	9.25	1.27
ㅑ	110C	98303	7.47	3.27	ㅕ	116E	99602	7.73	3.32	ㅑ	11A8	31752	7.99	1.10
ㅎ	1112	89829	6.83	2.99	ㅡ	1173	98447	7.64	3.28	ㅗ	11B7	29452	7.41	1.02
ㄴ	1102	72781	5.53	2.42	ㅞ	1162	45011	3.49	1.50	ㄴㅎ	11AD	13040	3.28	0.45
ㅗ	1106	64632	4.91	2.15	ㅋ	1167	33685	2.61	1.12	ㅎ	11C2	12992	3.27	0.45
ㅓ	1107	57483	4.37	1.91	ㅋ	1166	16696	1.30	0.56	ㅓㅏ	11B9	10072	2.53	0.35
ㄷ	1105	49339	3.75	1.64	ㅋ	1168	7802	0.61	0.26	ㅓ	11B8	8862	2.23	0.31
ㅕ	1104	23752	1.81	0.79	ㅓ	116D	6866	0.53	0.23	ㅕ	11C0	7785	1.96	0.27
ㅑ	110E	12917	0.98	0.43	ㅍ	1172	1764	0.14	0.06	ㅍ	11C1	4532	1.14	0.16
ㅍ	1111	7728	0.59	0.26	ㅑ	1163	1068	0.08	0.04	ㅑ	11AE	4159	1.05	0.14
ㅕ	110A	6720	0.51	0.22						ㅏ	11BA	3559	0.90	0.12
ㅕ	1110	5132	0.39	0.17						ㅑ	11BD	3268	0.82	0.11
ㅋ	110F	3862	0.29	0.13						ㄴㅑ	11AC	1029	0.26	0.04
ㅑ	1101	2757	0.21	0.09						ㅑ	11BE	1024	0.26	0.04
ㅓ	1108	2019	0.15	0.07						ㄷㅗ	11B1	1013	0.25	0.04
										ㄷㅑ	11B0	2098	0.23	0.03
										ㅋ	11BF	130	0.03	0.00
계		1315576		43.80	계		1289053		42.92	계		398833		13.28
총빈도수 : 3,003,462														

존 부호체계를 최대한 유지하면서 개선하는 방안이다. 표 8은 이것을 보여주고 있다.

표 8의 방안은 원천부호화에 위배되는 옛글자를 제외하고 초성과 종성의 “ㄱㄴㄷ” 및 “ㄱ”을 고려할 때 총 사용 빈도율은 19.78%가 된다. 즉, 이 방안을 채택할 경우 데이터 전송할 때 발생하는 스크램블링을 19.78% 떨어뜨릴 수 있다는 의미를 갖게 된다.

표 8. 기존 부호체계를 최대한 유지하는 방안
Table 8. Remaining of existing code system

열 \ 행	4	5	6	7
0				
1		표		·
2		ㅎ	ㄱ	△
3		△	ㄴ	ㅇ
4	ㄱ	ㅇ	ㄷ	ㅎ
5	ㄴ	ㅎ	ㄹ	\
6	ㄷ	ㅏ	ㅓ	/
7	ㄹ	ㅑ	ㅕ	:
8	ㅓ	ㅓ	ㅗ	.
9	ㅕ	ㅓ	ㅇ	
10	ㅗ	ㅓ	ㅗ	
11	ㅇ	ㅓ	ㅗ	
12	ㅗ	ㅓ	ㅗ	
13	ㅗ	ㅓ	ㅗ	
14	ㅗ	ㅓ	ㅗ	
15	ㅗ	ㅓ	ㅗ	

두 번째 방안은 표 9와 같이 기존 부호 배열 판 내에서 사용빈도에 따라 부호를 교체하는 방안이다. 예를 들어 상위비트 열 4의 경우에 사용빈도가 높은데 원천부호화 규칙에 위배되는 초성의 “ㄱㄴㄷ”의 부호는 사용빈도가 낮은 “ㅋㅌ표” 부호와 교체하는 것이다. 상위 비트 열 6의 경우에 사용빈도가 높은 종성의 “ㄱ”의 부호를 “△”부호와 교체하는 것이다. 상위비트

열 5와 7에 있는 위배 부호 “ㅇ”는 인접한 부호와 함께 위배가 일어나지 않는 부호로 옮기는 것이다. 이렇게 할 경우 원천부호화에 위배되는 낱자 “ㅋㅌ표”의 사용 빈도율은 합 1.39%가 된다. 즉, 전송효율 측면에서 볼 때 19.78%-1.39%=17.39%의 개선효과를 얻게 된다.

표 9. 기존 부호체계내에서 사용빈도에 따라 부호간 교체

Table 9. Alternation between codes with frequency in existing code system

열 \ 행	4	5	6	7
0	(초성)		(종성)	
1	ㅋ	ㅇ	△	ㅇ
2	ㅌ	ㅎ	ㄴ	ㅎ
3	표	ㅏ	ㄷ	\
4	ㄹ	ㅑ	ㄹ	/
5	ㅓ	ㅓ	ㅓ	:
6	ㅕ	ㅓ	ㅕ	.
7	ㅗ	ㅓ	ㅗ	
8	ㅇ	ㅓ	ㅇ	
9	ㅗ	ㅓ	ㅗ	
10	ㅗ	ㅓ	ㅗ	
11	ㅗ	ㅓ	ㅗ	
12	ㅗ	ㅓ	ㅗ	
13	ㅗ	ㅓ	ㅗ	
14	ㅗ	ㅓ	ㅗ	
15	ㅗ	ㅓ	ㅗ	

세 번째 방안은 상위비트 열 4번과 6번을 바꾸는 것이다. 즉 초성과 종성의 부호를 바꾸는 것이다. 이것은 초성이 종성보다 사용빈도가 높기 때문에 원천부호화 규칙에 위배되는 부호가 많은 상위 비트열 4를 상위비트열 6으로 옮겨 바꾸는 것이다. 이 방안의 경우에는 원천부호화에 위배되는 부호의 사용빈도율이 초성 “ㄱ” 종성 “ㄱㄴㄷ” 각각 5.89%, 1.1%, 2.98% 및 0.14%로 합 6.1%가 된다. 즉 전송효율 측면에서의 개선효과는 19.78%-6.1%=13.68%가 된다.

표 10. 사용빈도에 의해 초성과 종성의 부호열을 교체
Table 10. Alteration of code sequence of first and last vowel with frequency

열 \ 행	4	5	6	7
0	(중성)		(초성)	
1	ㄱ	ㅇ	ㄱ	ㅇ
2	ㄴ	ㅇ	ㄴ	ㅇ
3	ㄷ	ㅏ	ㄷ	\
4	ㄹ	ㅑ	ㄹ	/
5	ㅓ	ㅕ	ㅓ	:
6	ㅖ	ㅗ	ㅖ	.
7	ㅘ	ㅙ	ㅘ	
8	ㅛ	ㅜ	ㅛ	
9	ㅝ	ㅠ	ㅝ	
10	ㅞ	ㅡ	ㅞ	
11	ㅋ	ㅣ	ㅋ	
12	ㅌ	ㅚ	ㅌ	
13	ㅍ	ㅜ	ㅍ	
14	ㅎ		ㅎ	
15	△		△	

IV. 결 론

본 논문은 3 비트 열, 4비트 행, 7비트 1바이트 부호 체계를 갖는 한글정음부호체계에 대하여 논하였다. 연구결과 한글정음부호체계를 사용할 경우 원천부호화 규칙에 위배되는 부호의 총 사용 빈도율이 19.78%에 달하는 것으로 나타났다. 결과적으로 이 부호체계를 사용할 경우 원천부호화 규칙에 부합되는 부호체계를 사용할 때에 비하여 데이터의 전송효율을 19.78%만큼 떨어지게 한다. 그러므로 이에 대한 개선 방안을 제시한 결과 부호 열에서 초성, 중성 및 종성 간의 구분이 명확하지 않지만 데이터 전송효율측면에서 개선효과가 가장 높은 표 8의 방안을 한글정음부호체계의 개선방안으로 제시한다.

Reference

[1] Behrouz A. Forouzan, "Data communications" McGraw Hill Korea, pp132-134. Jan, 2008.
[2] <http://en.wikipedia.org/wiki>

/HDB3#B8ZS_.28North_American_T1.29
[3] Standard ECMA-35, "Character Code Structure and Extension Techniques", 6th Edition-December 1994, Standardizing Information and Communication Systems. Fig 7. p21.
[4] <http://en.wikipedia.org/wiki> /HDB3#HDB3_.28European_E-carrier.29
[5] <http://searchnetworking.techtarget.com/definition/B8ZS>
[6] http://en.wikipedia.org/wiki/Huffman_coding
[7] ITU-T, "Physical/electrical characteristics of hierarchical digital interfaces", ITU-T Recommendation G.703, Oct. 1998.
[8] TTA, "Test Method for Telecommunication Terminal Equipment", TTA Standard TTAS. KO-05.0028/R1, Revised on 23 Dec. 2004.
[9] W.P Hong "Study of ASCII Coding system considering of transmission efficiency in datacommunications" KIECS Journal, Vol.6 No5. 2011. 10.
[10] W.P Hong "Character coding rule of 1byte, 3x4bits considering the transmission efficiency in datacommunications" KIECS Journal, Vol.6 No4. 2011. 08.
[11] W.P Hong "Character coding rule of 2byte, 4x4bits considering the transmission efficiency in datacommunications", KONI Journal, Vol 15, No5, 2011. 10.
[12] <http://en.wikipedia.org/wiki/Unicode>
[13] W.P Hong "Datacommunications and Hangeul Jamo coding system of 7bits", KIMICS proceeding, Vol 1, pp246-255. 2012. 08.

홍 완 표 (洪完杓)



1991년 서울과학기술대학교 전자공학과 졸업(공학사)
1994년 연세대학교대학교 공학대학원 산업공학과 졸업(공학석사)
1999년 광운대학교 대학원 전자공학과 졸업(공학박사)
1990년 전기통신기술사합격
1991년 정보통신부 5급특별채용고시 합격 본부 통신정책실, 전파방송관리국, 정보화기획실
1997년 삼성전자(주) 통신사업부 전송영업그룹장
1999년 광운대학교 연구전담교수
2000년 한국정보통신기술협회장
2002년 한세대학교 IT학부 정보통신공학전공 교수
한세대학교 정보통신연구소장
관심분야 : 위성통신방송/문자코딩/통신정책/