

論文2002-39CI-5-4

TV 자막 신호를 이용한 한글 수화 발생 시스템의 개발

(Development of Korean Sign Language Generation System using TV Caption Signal)

金大珍*, 金亭培*, 張源**, 卞增男*

(Dae-Jin Kim, Jung-Bae Kim, Won Jang, and Zeungnam Bien)

요 약

본 논문에서는 TV 자막 신호를 이용한 한글 수화 발생 시스템을 다룬다. TV 자막 방송 프로그램에 포함된 자막 신호는 자막 신호 복호화기(Caption Decoder)를 이용하여 PC로 전송된 후, 한글 수화의 특성에 적합하게 개발된 형태소 분석기를 통하여 의미 있는 단위로 나누어 진다. 분석된 형태소는 3차원 수화 애니메이션 데이터로 변환되어 3차원 한글 수화 발생기를 통하여 시각적으로 표현된다. 특히, 실시간 처리가 가능하도록 각종 전처리 기법들에 기반한 형태소 분석기를 제안하였다. 개발된 시스템은 실제 자막 방송 프로그램에 적용되어 그 유용성을 검증하였으며 실제 농아인들의 사용에 의하여 그 실용성을 검증하였다.

Abstract

In this paper, we propose TV caption-based KSL(Korean Sign Language) generation system. Through TV caption decoder, this caption signal is transmitted to PC. Next, caption signal is segmented into meaningful units by morphological analyzer in considering specific characteristics of Korean sign language. Finally, 3D KSL generation system represents the transformed morphological information by 3D visual graphics. Specifically, we propose a morphological analyzer with many pre-processing techniques for real-time capability. Our developed system is applied to real TV caption program. Through usage of the deaf, we conclude that our system is sufficiently usable compared to conventional TV caption program.

Key Words: KSL(Korean Sign Language) Generation(한글 수화 발생), TV Caption(TV 자막), Morphological Analysis(형태소 분석)

I. 서 론

자막 방송이란 TV방송 전파의 송신단에서 TV신호의 VBI(Vertical Blanking Interval)구간에 TV방송 프로그램의 음성신호와 동기를 이루는 자막 신호를 부호화하여 전파를 통해 송신한 후, 이를 자막 방송 수신부에서 복호화하여 시청자가 이러한 자막을 선택적으로 볼 수 있도록 하는 TV 부가 방송 서비스의 일종이다.^[1] 현재 국내에서는 1999년 2월 12일 최초로 한글 자막 방송을 시작한 이래, 각 방송사에서 3월경부터 자막 방

* 正會員, 韓國科學技術院 電子電算學科 電氣 및 電子 工學

(Division of Electrical Engineering, Department of Electrical Engineering and Computer Science, KAIST)

** 正會員, 國防科學研究所

ADD(Agency for Defense Development)

※ 본 연구는 (주) 삼성전자, 인간친화 복지 로봇시스템 연구센터로 부분적 재정을 지원받아 수행되었다.

接受日字:2001年8月31日, 수정완료일:2002年8月19日

송을 개시하여 현재 운영 중에 있으나 대상 방송 프로그램이 한정되어 일 2~3시간 분량의 방송 프로그램들에 한해서만 자막 방송이 적용되고 있다. 1998년 5월 8일에는 정보 통신부의 인가를 받은 사단법인 한국자막 방송기술협회가 정식 발족되어 자막 서비스를 각 방송사에 제공하고 있다. 미국의 경우 4대 공영 방송이 모든 방송을 자막 방송 처리하고 있으며, 유럽, 남미 등지에서도 1970년대부터 자막 방송을 시행해오고 있고, 방송시간 또한 주당 30~50시간으로 우리 나라와 큰 차이를 보인다. 미국의 경우 자막 방송 전달기구인 NCI(National Captioning Institute)가, 캐나다의 경우에는 CRC(Communications Research Centre)가 자막 방송에 필요한 지원을 하고 있다. 전 세계의 TV방송 기술이 디지털 기반으로 이행하면서 디지털 캡션 기술을 이용한 자막 방송이 전 세계적으로 보편화될 전망이다.

TV 자막방송 잠점표준에 의하면, 자막방송의 취지는 청각장애인의 복지 증진 및 교육에 활용하며 유용한 부가 정보를 제공하는데 있다.^[2] 현재 국내의 자막방송 수혜자는 청각장애인 40만명을 포함해 난청환자, 노인 등이 100만에 이르고 한글의 조기 교육, 외국인의 한글 학습 등의 기타 이유로 자막 방송을 필요로 하는 인구가 수백만에 이를 것으로 추산된다.^[3]

기존의 자막 방송 프로그램 혹은 청각장애인을 위한 수화 통역사를 동반한 방송 프로그램의 경우, 현실적으로 다양한 문제점을 지니고 있다. 자막 방송의 경우에, 혜택을 받는 농아인의 수는 전체 농아인들 중 4분의 1도 되지 못한다. 여전히 대부분의 농아인들이 수화 표현에 더 애착을 가지며 이해도가 높다고 한다. 그러나 수화 통역사의 수도 매우 제한적이어서, 방송 프로그램을 위한 수화 통역사 확보에 많은 어려움을 겪고 있다. 결국, 방송 프로그램의 수나 시간에 있어서 매우 제한적인 수밖에 없다.

자막 방송의 시청을 위해서는 디코더(Decoder)가 필요하며 현재 일부 업체에서 디코더 내장형 TV 수신기를 개발 및 양산 중에 있으나, 대상 TV 수신기가 29인치 이상의 대형이어서 고가이며, 상대적으로 한글 자막 방송을 필요로 하는 대상자들의 경제 수준이 높지 않아 거의 실효를 거두지 못하고 있는 상태이다.

한국과학기술원 전자전산학과에서는 3차원 애니메이션을 이용한 한글 수화 발생기를 개발하였다.^[4] 수화

교육용 도서 및 수화 사진 등을 참고하여 구축한 수화 DB에 근거하여, 키프레임(Key-frame) 방식으로 3차원 수화 애니메이션을 발생시키는 한글 수화 발생기는 자기 센서와 사이버 글러브(CyberGloveTM) 등으로 한글 수화 인식 시스템과 함께 한글 수화 통역 시스템을 구성한다. 한글 수화 발생기는 자체적으로 입력된 국어 문장을 형태소 분석(Morphological Analysis) 과정을 통하여 수화 단어로 변형하고, 변형된 수화 단어에 해당하는 3차원 애니메이션을 표현함으로써 농아인들의 이해를 돕는 역할을 한다.

본 논문에서 기술하는 자막방송 기반 한글수화 발생기는 기존의 자막 방송 프로그램 및 수화 통역사에 의한 방송 프로그램의 장점을 혼합한 형태라고 할 수 있다. 즉, 자막 방송의 자막 신호를 한글 수화 발생기의 입력으로 사용하는 것이 자막 방송 기반 한글 수화 발생기의 근본적인 개념이다.

본 논문의 구성은 다음과 같다. 우선, II장에서는 자막 방송 신호를 수화로 변환하기 위한 형태소 분석기에 대하여 다룬다. III장에서는 형태소 분석기의 결과로 얻어진 수화를 3차원 애니메이션으로 표현하는 한글 수화 발생기의 구성에 대하여 설명하고, IV장에서는 자막 방송 기반 한글 수화 발생기의 전체 구성 및 응용에 대하여 기술한다. 마지막으로 V장에서는 결론과 추후 과제를 제시한다.

II. 자막 신호 변환을 위한 형태소 분석기

자막 신호는 다수의 타자수(Typist)에 의하여 입력된 자막 데이터를 TV 방송 신호에 부호화(Coding)한 형태로 전송된다. 일반적으로 전송되는 자막 방송 데이터는 농아인들을 위한 표현이라기 보다는 정상인들이 일상 생활에서 사용되는 표현들이 대부분이다. 따라서, 자막 신호를 수화식 표현으로 변환하는 작업이 필요한데 이를 위해서 기존의 형태소 분석기에 대한 연구결과를 활용할 수 있다.

형태소 분석기는 한글 문장 혹은 어절을 더 작은 단위의 형태소로 분류하는 기능을 한다. 형태소 분석기와 관련된 연구 중에서 가장 대표적인 것은 한국과학기술원 전자전산학과에서 개발 중인 한글 태깅(Tagging) 시스템을 들 수 있다.^[5] 현재 약 8만여개의 단어 DB를 갖고 있으며, 불규칙 변화 및 동사변형, 생략, 준말, 맞춤법에 관련된 여러 가지 규칙들을 프로그래밍 기법을

1) 최근에는 '농아인'이라는 표현이 표준어로 사용된다.

통해 처리한다. 실제 형태소 분석에 있어서는 확률적 방법에 의존하는 HMM (Hidden Markov Model) 기법을 사용하고 있다. 1Mbyte (한글 약 50만자) 에 해당하는 파일을 형태소 분석하는데 약 1분 정도가 소요되며, 전반적인 성공률은 90% 정도이다. 하지만, 실시간으로 자막 방송 신호를 처리하기에는 그 속도가 충분하지 않다.

국어의 특성상 용언 표현에 있어서의 불규칙 변화는 반드시 고려해야 할 대상이다. 용언에 대한 불규칙 변화를 처리하기 위해서는 현재 분석된 형태소에 대한 품사 정보가 부가적으로 필요하게 된다. 이를 위하여 본 논문에서는 좌우접속정보 코드라는 것을 이용하고 있는데 여기에는 각 형태소의 품사 정보 및 연결 정보 등이 포함되어 있다.^[7] 이러한 품사 정보를 이용하면, 불규칙 변화 이외에도 다양한 규칙들을 구성하고 효율적인 수화 변환이 가능하게 된다. 수화 표현을 위해 불필요한 조사, 접사, 어미 등을 생략하는 데에도 이러한 품사 정보를 유용하게 사용할 수 있다. 실제로 수화 통역사로서 방송 프로그램에 참여하는 분들에 의하면, 방송 중에는 그때마다의 적절한 판단에 따라 수화를 적당히 생략하기도 하고 좀 더 쉬운 말로 표현하기도 하는 등 원래 방송 대본에 많은 수정을 가한다고 한다.

한글 문장 및 어절을 처리하기 위한 형태소 분석기는 원래 SGI 워크스테이션에서 운용되던 것을 PC의 윈도우(Windows) 환경으로 포팅(Porting)하여 사용하고 있다.^[4] 수화 단어 DB를 검색하는데 있어서는 가장 기본적인 순차검색(Sequential Searching)의 형태를 지니고 있다. 순차검색의 경우, 데이터의 수가 증가할수록 검색 속도 및 효율이 현저히 감소하는 것으로 알려져 있으나, 최근 컴퓨터 및 관련 기술의 발달로 속도가 저하가 문제된 경우는 없었다.

이상의 관찰을 바탕으로, 본 논문에서는 다음과 같은 점에 중점을 두어 자막 신호 변환을 위한 형태소 분석기를 구성하였다.

- 1) 실시간 시스템 구현을 위한 불필요한 조사 및 어미의 생략
- 2) 과거형에 대한 전처리
- 3) 고유명사에 대한 전처리
- 4) 불규칙 활용에 대한 전처리
- 5) 복수형에 대한 전처리

이러한 점들은 수화의 특성상 과거라든가 미래에 대

한 시제 개념이 없고, 수동태/능동태의 구분이 없으며 특별한 조사 없이도 의미의 구성이 쉬운 점들에 착안하였다. 또한, 불규칙 활용 문제 역시 앞서 지적된 것처럼 형태소 분석기 차원에서 해결되어야 할 큰 문제 중의 하나이다. 고유명사는 대부분 특정인의 이름(방송중의 등장인물이나 유명인사 등)을 지칭하는 경우가 많음에 착안하여, 지화로 표현하지 않고 그에 해당하는 그림 등으로 간략히 표시하면 더욱 효과적일 것이다. 이에 적합한 수화 DB의 구성도 필요하게 되는데, 여기에는 품사 정보가 포함된다. 이하에서는 개발된 형태소 분석기 전반에 관한 기술과 수화 DB의 구조, 그 외의 각종 전처리 기법들에 관하여 상세히 기술하기로 한다.

(1) 형태소 분석기 전체 시스템 구성

전체 형태소 분석기는 그림 1과 같은 흐름도로 구성되어 있다.

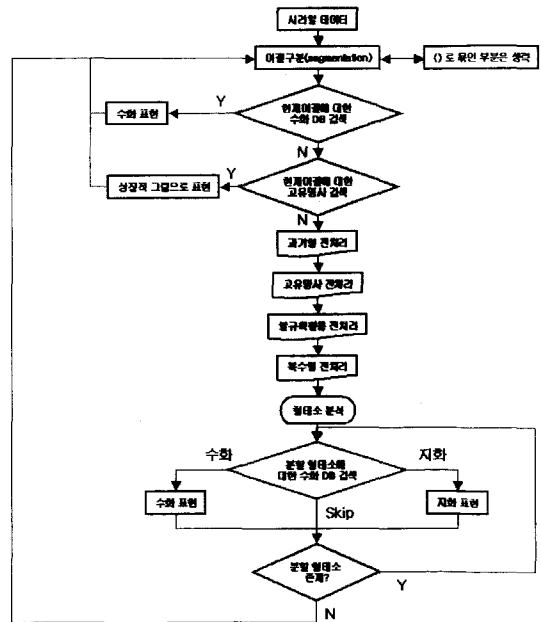


그림 1. 제안된 형태소 분석기의 흐름도
Fig. 1. Flow chart of proposed morphological analyzer.

막 데이터가 입력되면, 어절구분 과정을 통해 실제 형태소 분석에 사용하기 위하여 어절 단위로 구분된다.^[4] 어절 구분은 공백 및 각종 문장 부호 등에 해당하며, 이 중에서도 괄호 (‘ 나 ‘) 는 대본에서 지문 처리에 종종 사용되므로 괄호로 묶인 문장 혹은 단어 전체를 생략하게 된다.

구분된 어절은 우선적으로 수화 DB를 검색하여 일치하는 경우 해당 수화를 표시하고, 고유명사가 존재하는 지도 검사한다. 다음으로는, 앞서 설명한 여러 가지 전처리(과거형, 고유명사, 불규칙활용, 복수형 등에 대한)를 행하고 이 결과는 기존의 형태소 분석기를 통하여 여러 개의 형태소로 분석된다.

마지막으로, 분할된 형태소 각각에 대하여 수화 DB 검색을 통하여 수화/지화로 구분 및 표현을 수행하고 생략(Skip)해야 할 형태소의 경우에는 생략 처리를 한다.

이러한 처리는 한 어절에 대한 분할 형태소 모두에 대한 처리가 끝날 때까지 지속되며, 전체 구성에서는 더 이상 처리할 자막 데이터가 없는 경우에 종료된다.

(2) 실시간 시스템 구현을 위한 자막 신호 처리

실시간 시스템 구현을 위하여 불필요한 정보를 생략하는 것은 필수적이다. 실제 표현에 있어서 조사·어미·접사·시제 등을 표현하게 되면 실시간 번역이 어려워지고, 표현되는 양이 많아서 청각장애인들에게 상당한 부담을 가져오게 된다. 통계치에 의하면, 능숙한 성인 수화 화자의 평균 지문자 표현 속도는 1분당 300문자라고 한다.^[6] 또한, 수화와 음성언어의 발신 소요 시간을 비교해 본 결과에 따르면 수화가 약 1.42배나 더 걸린다는 보고도 있다.

본 논문에서는 이러한 문제점을 해결하기 위하여 좌우접속정보 코드를 이용하여 분석된 형태소의 품사정보를 추출함으로써 어미 및 조사, 수식언어에 대한 생략 기능을 구현하였다. 좌우접속정보 코드는 16진수 6자리로 구성되어 있으며, (각 자리를 편의상 "ABCDEF"로 표기) 형태소 해석용 접속정보를 저장하고 있다. 이때, 품사코드는 정보코드의 앞쪽 2자리 코드인 "AB"를 지칭하며 좌정보코드는 앞쪽 4자리 코드 "ABCD", 우정보코드는 가운데 2자리를 제외한 코드인 "ABEF"를 지칭한다.^[7]

생략 부분은 비교적 간단하게 구성되어 있다. 우선 어절에서 분할된 형태소에 대한 접속정보를 읽고, 조사(4xxxx) 혹은 종결어미(91xxxx)인 경우 생략 정보를 추가한다. 아래의 예는 실제 자막 방송(드라마 중의 한 부분)에서 추출된 자막 데이터에 생략 기능을 적용한 결과이다²⁾.

2) 수화 단어는 (수화), 지화 단어는 (지화), 생략 단어는 (Skip) 으로 표기됨.

축하해 → 축하(수화) + 해(Skip)

또한, 생략 기능의 신뢰성 확보를 위하여 현행 고교 문법 교과서 및 웹 검색 자료를 기준으로 국어에서 사용되는 대표적인 어미들을 정리하는 작업을 선행하였다. 현재 총 246개의 어미가 수화 DB에 수록되어 있으며, 이것은 고교 문법 교과서 및 웹을 통해 검색된 대부분의 어미를 포함한다.

(3) 수화 DB의 구성

품사정보를 이용한 형태소 분석을 위해서 수화 DB는 다음과 같이 구성된다.^[4] 수화 DB에서 한 단어에 대한 수화 데이터는 해당 수화에 대한 문자열 데이터와 수화 동작을 위한 전체 키프레임(Keyframe) 수³⁾, 해당 단어가 갖는 품사의 수, 품사 코드, 각 키프레임을 표현하기 위한 지연시간(Delay) 및 표정코드, 56개의 3차원 애니메이션 데이터의 조합으로 이루어져 있다.

수화단어	키프레임수	품사수	품사코드
(1번째 키프레임)	지연시간	표정코드	56개의 3차원 애니메이션 데이터
(2번째 키프레임)	지연시간	표정코드	56개의 3차원 애니메이션 데이터
...			

이러한 과정을 통해 구축된 수화 DB는 현재 순수 단어(품사 정보를 고려하지 않은 형태소)가 1053개, 품사 정보를 고려한 모든 단어의 경우 1211개를 포함하고 있다.

(4) 과거형에 대한 전처리

과거형에 대한 전처리를 수행하기 위한 흐름도는 그림 2에 나타나 있다.

우선, 입력된 하나의 어절을 음절 단위로 분리한다. 한글 자막에 대한 표준 규약을 참조하면, 현재 자막 방송에 사용되는 한글 자막은 KSC-5601 완성형 코드를 이용하고 있다.^[2] 따라서, 이를 3바이트 초/중/종성 코드로 변환하면 유용하게 활용할 수 있다. 일반적으로 과거형은 종성에 'ㅅ'이 포함되므로, 본 논문에서 제안하는 방법도 이를 기준으로 과거형을 식별한다. 용언에 따라서는 원형이 과거형으로 변형될 경우, 단모음이 복모음으로 표현되는 경우가 있다. 따라서, 이들 각각에 대한 처리를 그림 2에서와 같이 행한다. 하지만, '있다'

3) 모델 구성 및 56개의 3차원 애니메이션 데이터는 III장에서 자세히 기술한다.

의 경우와 같이 과거형이 아닌데도 원형에서 'ㅅ'이 포함되는 경우도 있으므로 이에 대한 예외처리도 포함되어 있다. 그 외에도 '있', '엿', '였' 등의 과거형 선어말 어미에 대한 처리도 가능하다. 또, '겠'과 같이 추측을 나타내는 선어말 어미에 대한 예외 처리도 포함하였다. 다음은 제안된 방법을 실제 자막 방송 데이터에 적용해 본 결과이다.

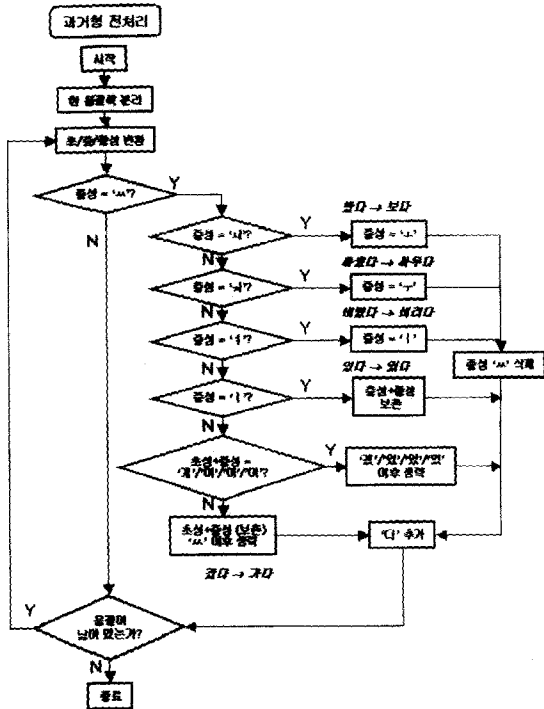


그림 2. 과거형에 대한 전처리 흐름도
Fig. 2. Flow chart for pre-processing of the past.

다녀왔대요 → 다녀오(수화) + 다(Skip)

(5) 고유명사에 대한 전처리

고유명사에 대한 전처리는 고유명사를 포함한 어절이 입력된 경우, 고유명사를 제외한 형태소를 삭제하고 최종적으로 고유명사만 포함한 어절로 변환하는 과정을 의미한다. 즉, 고유명사의 경우 그에 해당하는 대표적인 (상징적인) 그림만으로 그 의미를 충분히 전달할 수 있으므로 나머지 조사나 어미 등을 생략하는 것이다. 그림 3은 고유명사에 대한 전처리 흐름도를 나타내고 있다.

여기서, 고유명사 DB는 현재 자막 방송에서 등장하는 인물이나 대표적인 고유명사들에 대한 내용을 포함

하고 있다. 실제 자막 방송에 제안한 방법을 적용한 결과는 다음과 같다. 단, 여기서는 분류의 편의상 고유명사를 '수화'로 표기하였음을 밝히는 바이다.

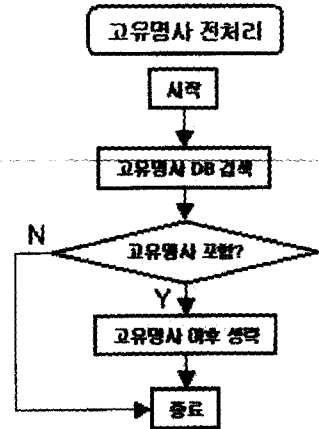


그림 3. 고유명사에 대한 전처리 흐름도
Fig. 3. Flow chart for pre-processing of a proper noun.

씨씨의 → 씨씨(수화) + 의(Skip) → 씨씨(수화; 고유명사)

(6) 불규칙 활용에 대한 전처리

불규칙 활용의 경우에도 고유명사의 경우와 마찬가지로 불규칙활용 DB가 요구된다. 본 논문에서는 불규칙 활용 DB 구성을 위하여 연세한국어사전을 참조하였다.^[8] 입력된 어절 중에서 불규칙 활용 DB에 포함된 활

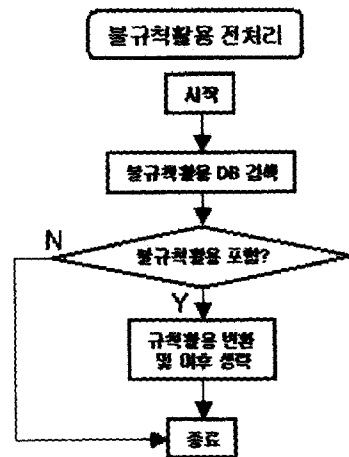


그림 4. 불규칙 활용에 대한 전처리 흐름도
Fig. 4. Flow chart for pre-processing of irregular conjugation.

용형이 발견될 경우, 이를 불규칙활용의 기본이 되는 원형으로 다시 변환하는 것이 이 방법의 핵심이다. 그림 4는 불규칙 활용에 대한 전처리 흐름도를 나타내고 있다.

다음은 불규칙 활용 DB의 한가지 예와 실제 자막 방송 데이터에 대하여 제안된 방법을 적용한 결과이다.

부르다 4 부르는 부른 불려 부릅니다

불려 → 부르다 → 부르(수화) + 다(Skip)

(7) 복수형에 대한 전처리

복수형의 경우, '들'과 같은 복수형 접미사를 포함하는 경우가 대부분이다. 따라서, 이들에 대한 전처리는 '들'이 현재 입력된 어절 내의 최종 위치에 존재하는지 검토함으로써 수행된다. 단, '들'이 포함된 경우에도 예외의 경우가 있으므로 복수형 예외 DB를 참조하여 전처리를 행한다. 실제 복수형에 대한 전처리 흐름도는 그림 5에 나타나 있다.

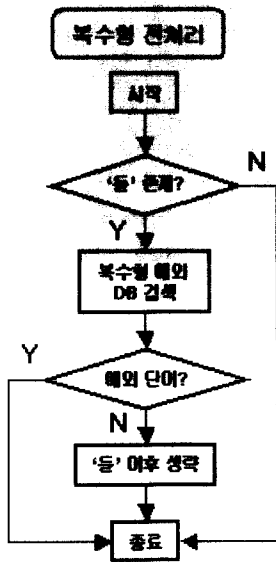


그림 5. 복수형에 대한 전처리 흐름도
Fig. 5. Flow chart for pre-processing of the plural form.

다음은 실제 자막 방송에 대하여 제안된 방법을 적용한 결과이다.

4) '들', '아들' 등이 여기에 해당된다.

친구들 → 친구(수화) + 들(복수형) → 친구(수화)

이러한 다양한 전처리 방법을 적용하여, 제안된 방법은 원래의 자막 방송 신호에서부터 수화를 표현하기 위해 필요한 핵심적인 자막 신호들을 추출할 수 있다. 즉, 주어진 자막 신호 문장에서 수화 표현에 불필요한 부분을 생략하고 한글 표준 수화의 형태로 변형함으로써 농이인들로 하여금 자막 방송 프로그램에 대한 이해도를 높이는 역할을 하게 된다.

III. 한글 수화 발생기의 구성

본 논문에서 다루고 있는 한글 수화 발생기에서는 3차원 수화 애니메이션 표현을 위해 대표적인 3차원 프로그래밍 도구로써 OpenInventor™를 이용한다.^[4] 3차원 캐릭터는 몸통, 머리, 손 및 팔 등으로 모델링되며, 형태소 분석기의 결과로 얻어진 3차원 수화 애니메이션 데이터에 기준하여 위치 및 방향 등이 갱신되어 표현된다. 또한, 각 수화 단어에 포함된 표정 코드를 기반으로 하여 해당 얼굴 표정을 나타낸다. 그림 6은 한글 수화 발생기의 기본적인 구성을 보여준다.



그림 6. 한글 수화 발생기의 기본 구성
Fig. 6. Basic Configuration of KSL Generator.

그림 6에서 나타난 것처럼 한글 수화 발생기는 한글 수화를 표현하기 위한 수화 표현자(Sign Language Generator)로써 3차원 캐릭터를 이용하고 있으며, 손 및 팔의 동작에 의해 지화 혹은 수화를 표현한다. 또한, 자막 방송을 위해 그림 6의 기본적인 한글 수화 발생기에 첨가하여 몇 가지 새로운 요소가 추가되었는데

여기에 대해서는 뒤에서 다루기로 한다. 본 절에서는 한글 수화 발생기의 화면 구성 및 3차원 캐릭터에 관해 상세히 기술하기로 한다.

1. 3차원 캐릭터 모델링

3차원 캐릭터의 모델링을 위해서 사람을 구성하는 각 모델 부위간의 우선 순위를 그림 7에 도시했다. 즉, 그림 7은 사람을 표현하기 위해서 각 부분 물체간의 모델을 구조적인 계층적 관계로 도시한 것이다.

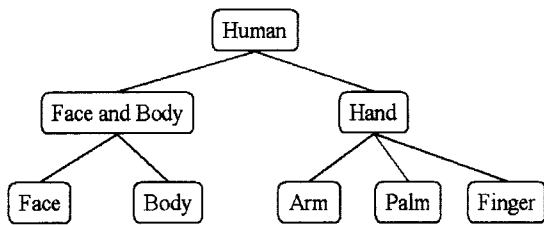


그림 7. 사람 모델에 관한 계층적 구조
Fig. 7. Hierarchical structure of the human model.

한글 수화 발생기에서 중요한 부분은 양손, 몸, 얼굴로서 특히, 수화의 표현에는 양손의 움직임이 절대적으로 중요하다. 수화를 표현하는 수화 표현자를 화면에서 구현할 때는 손의 제스처를 가장 정확히 나타내어야 사람이 이해하기 쉽다. 또한, 농아인들의 경우 얼굴표정을 정상인에 비해 매우 강하게 표현하고 있으며, 의사소통 채널로써 유용하게 사용한다. 이를 표현하기 위해서는 얼굴 표정 모델의 추가가 필요하다.

이러한 관점에서 그림 7의 모델링 방법은 손의 움직임만 3차원으로 처리하고 나머지 부위는 2차원 형태로 간략히 처리할 수 있다는 장점을 갖는다. 또한, 전체 시스템 동작 시간에서 그래픽 표현을 위해 소요되는 시간을 효과적으로 줄일 수 있다. 실제 모델링에 있어서는 구(Sphere)와 기둥(Pillar) 모델을 사용하고 있다.⁵⁾ 구는 머리를, 나머지 부분인 몸통, 팔, 손, 각 손가락 등은 기둥 모델을 기초로 하여 모델링하였다.

그림 8은 한글 수화 발생기를 구현하기 위해 필요한 기본적인 요소들을 표시하고 있다. 여기서 몸, 얼굴은 앞서 설명한 바와 같이 2차원 형태로 간략히 표현하며

5) 이것은 얼굴 및 몸통의 모델이 2차원으로 구성되어 있다는 것이 아니라, 화면 구성에 있어서 2차원 위치 정보 만으로 간단히 표현할 수 있음을 의미한다.

손과 팔은 비교적 정교하게 모델링한다.

사람의 손을 표현하려면 자세하고 섬세한 모델링이 필요하지만, 전체 시스템의 실시간 동작을 보장할 수 있을 정도로 효율적이기도 해야 한다. 본 논문에서는 표현에 의한 3차원 모델링 기법을 이용하여 손 모델을 구성하였다 (그림 8).

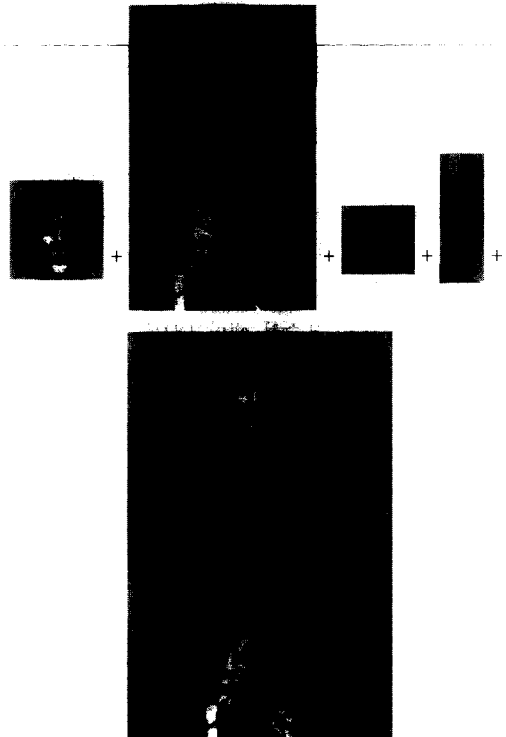


그림 8. 사람 모델을 위한 구성도
Fig. 8. Configuration of the human model.

손 모델은 손 본체의 모델과 손가락의 모델로 구성 되어 있으며, 각 손가락은 실제 사람의 손가락 마디와 동일한 개수로 구성되어 있다. 이 경우, 손가락이 펴진 상태에서는 특별한 문제점이 없으나 손가락을 구부린 형태를 표현하기 위해서는 간단한 좌표변환을 통하여 손가락 굽힘시 나타나는 모델간의 겹침 현상을 제거한다.¹⁰⁾

팔은 세 개의 관절을 가지고 있다. 사람의 어깨 관절은 3자유도, 팔꿈치는 1자유도, 손목은 3자유도 모델로 각각 표현 가능하다. 팔의 각 관절에 대한 자유도 표현은 3차원 회전(Roll-Pitch-Yaw) 표현을 이용하고 있다. 시간에 따라 연속적으로 움직이는 물체의 표현은 3차원 애니메이션에 있어서 매우 중요하다. 여러 가지 다

양한 방법들 중에서, 본 논문에서 다루고 있는 한글 수화 발생기는 애니메이션을 단순화시킨 키-프레임(Key frame) 방식을 사용한다. 여기서 '단순화'란 의미는 손의 위치와 방향 및 손가락의 회전량과 시간만을 프레임 데이터로 갖고 애니메이션을 수행하는 것을 의미한다. 수화의 한 정지 동작을 나타내는 데이터, 즉, 각 손의 위치와 방향, 각 손가락의 회전각과 다음 정지 동작을 표시하는 데까지의 시간을 데이터로 하여 수화 애니메이션을 표현한다. 앞서 III장에서 기술한 수화 DB에 있어서, 56개의 데이터는 이러한 키-프레임으로써 수화 애니메이션에 직접적으로 사용된다. 총 56개의 데이터는 각각 28개씩 양쪽 팔 및 손의 데이터에 해당된다. 즉, 전반부의 28개 데이터는 오른팔 부분에서 어깨, 팔꿈치, 손목의 3차원 회전(Roll, Pitch, Yaw) 정보와 각 손가락 마디의 굽힘 정보로 구성되어 있다. 후반부 역시 마찬가지로 왼팔 부분에 해당된다. 또한, 수화는 연속적인 동작으로 표현되므로 56개의 데이터 이외에도 지연 시간에 대한 부분이 추가되어 수화 애니메이션의 속도를 조정하는 역할을 한다. 지화 DB의 경우에는 31가지의 지화에 대하여 각각 정지 동작을 표현하기 위한 56개의 3차원 그래픽 데이터로 구성되어 있다. 또한, 초성·중성·종성을 표현하기 위해서 각 지화의 위치를 형태소 분석기의 결과에 따라 서로 다른 위치에 표현할 수 있는 기능도 구현하였다.

2. 자막 방송을 위한 한글 수화 발생기 화면 구성

자막 방송을 위한 수화 발생기를 구현하기 위해서는 사용자 친숙성이 가장 중요한 요소이다. 본 논문에서는 자막 방송을 위한 한글 수화 발생기를 구성하는데 있어서 다음과 같은 점들에 중점을 두었다.

- 1) 화자 변경에 따른 다양한 얼굴 모델을 사용하여 시각적 효과를 높인다.
- 2) 다양한 표정의 얼굴 모델을 이용하여 좀 더 자연스러운 표현을 가능하게 한다.
- 3) 고유 명사가 입력되면 해당하는 캐릭터 혹은 상징적 그림을 사용하여 표현한다.

화자 변경에 따른 얼굴 모델을 구현하기 위해 3차원 모델링 툴을 이용하여 다수의 얼굴 모델을 만들었다. 그림 9는 제작된 얼굴 모델 중에서 대표적인 몇 가지를 보인 것이다.

또한, 3차원 모델링 툴을 이용하여 그림 10과 같은 다양한 얼굴 표정 모델을 제작하고 수화 DB의 표정 코드에 따라 다양한 표정을 표현한다. 이것은 III-2절에서 기술된 바와 마찬가지로 얼굴표정이 농아인들의 의사소통에 있어 매우 중요한 수단으로 사용됨에 기인한다.

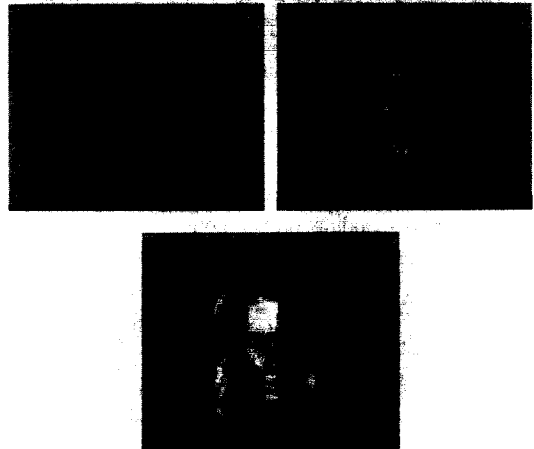


그림 9. 화자에 따른 다양한 얼굴 모델
Fig. 9. Various face models for different speakers.

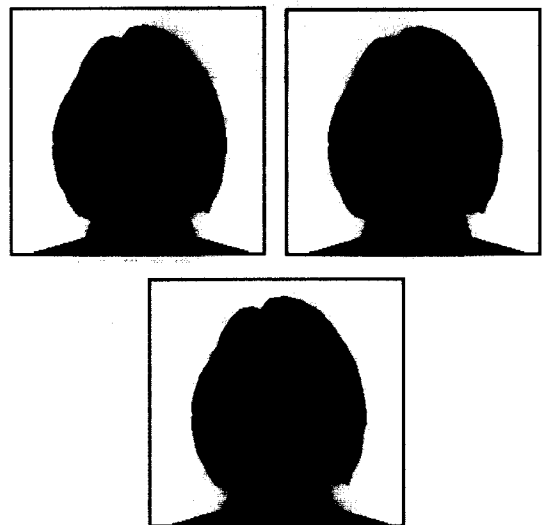


그림 10. 표정에 따른 다양한 얼굴 모델
Fig. 10. Various face models for different facial expressions.

II장에서 설명된 바와 같이, 고유 명사에 해당하는 캐릭터 혹은 상징적 그림 등은 굳이 수화나 지화로 표현하지 않고도 그 의미를 충분히 전달 가능하므로 이를 표현하기 위하여 3차원 캐릭터의 우상단에 소형 패



그림 11. 고유 명사에 해당하는 캐릭터를 표현하고 있는 소형 패널

Fig. 11. Mini panel for expressing the character of corresponding proper noun

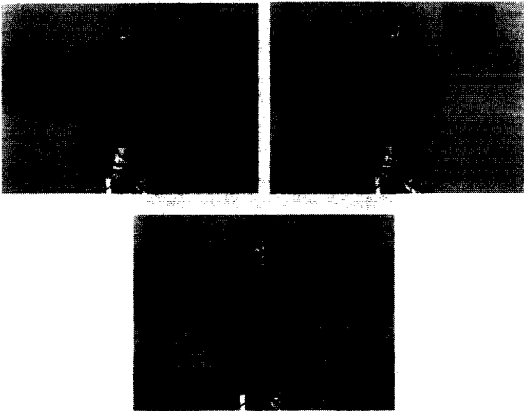


그림 12. 자막 정보가 순차적으로 표시되는 과정
Fig. 12. A sequence of KSL generation according to caption signal.

널(Panel)을 구성하였다. II장에서 설명된 고유 명사에 대한 전처리를 통하여 이 소형 패널에는 해당하는 그림이 표시된다. 그림 11은 실제 자막 방송에 적용하여 표현되는 고유 명사 패널을 보여 주고 있다.

그림 12는 실제 자막 방송 데이터를 이용하여 수화 발생기를 수행하는 한 가지 예를 보여 주고 있다. 이해를 돕기 위하여 하단에는 자막 정보를 동시에 표현하고 있다.

IV. 시스템 구성 및 운용

그림 13은 개발된 전체 시스템의 구성을 보여주고 있다.

TV 방송국에서 전달된 자막 신호를 자막 신호 디코더(Caption Decoder)를 통하여 추출하고 이것을 PC로 전달한다. PC에서는 입력된 자막 신호에 대한 전처리(Preprocessing) 및 개발된 형태소 분석기를 통하여 형태소 분석을 행한다. 형태소 분석의 결과는 미리 입력된 수화 및 지화 DB와 비교되어, 해당하는 3차원 그래픽 데이터 및 얼굴 표정 데이터를 얻게 된다. 이렇게 얻어진 3차원 데이터를 이용하여 개발된 수화발생기는 사용자에게 3차원 그래픽으로 구성된 애니메이션으로 보여주게 된다.

II장 및 III장에서 제안된 구성 요소들의 결합을 통해 자막 방송을 위한 한글 수화 발생 시스템을 구성하

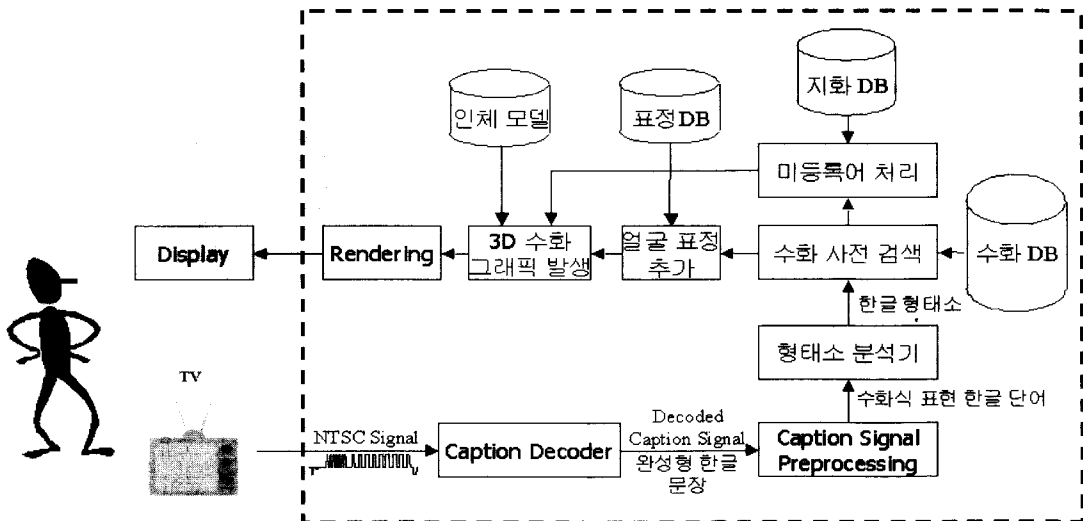


그림 13. TV 자막방송 기반 한글 수화 발생기의 구성
Fig. 13. Overall configuration of TV caption-based KSL generator.

고, 실제 자막 방송 프로그램 중 일부에 적용해 보았다. 다양한 실험에 의해, 제안된 방법은 원 문장의 이해도를 유지하면서도 수행 속도 면에서 효과적인 것으로 판단되었다. 실제로 이러한 전처리 방법을 이용하지 않을 경우, 자막 방송 프로그램을 한글 수화로 표현하는데 있어 시간 소모가 상당한 것으로 파악되었다 [4]. 본 논문에서는 현재 방영중인 다양한 자막 방송 프로그램 중 하나를 선정하여 제안된 시스템을 적용한 결과를 소개한다.

1. 자막 방송 프로그램에 대한 시스템 적용 결과

시스템 적용을 한 결과, 제안된 방법을 이용한 경우 생략된 단어들을 제외한 전체 단어 중에서 약 80%에 해당하는 단어들이 수화로 표현됨을 알 수 있었다.

표 1은 개발된 시스템을 이용하여 자막 방송 프로그램에 적용한 결과를 나타낸다. 참고로, 본 프로그램의 전체 방영 시간은 2분 40초 정도이며 표 1에 나타난 수행시간은 순수하게 수화 발생기에서 3차원 그래픽 처리에 소요된 시간을 의미한다⁶.

표 1. 자막 방송 프로그램에 대한 시스템 적용 결과

Table 1. Results of applying developed system to one of the caption programs

	단어수 (전체비율)	frame 수	수행시간 (sec)	Frame /sec
수화	173 (약 50%)	444	39.43	11.2605
지화	59 (약 17%)	314	26.265	11.9551
전체단어 (skip 제외)	232 (약 67%)	758	65.695	11.5382

지화 단어의 비율이 전체 단어 중에서 대략 25% 정도를 차지하는데 반하여, 수행 시간에 있어서는 약 40%를 차지하고 있음을 알 수 있다. 지화 단어의 표현은 각 음절에 대하여 초/중/종성을 모두 표현하므로 적게는 2 프레임에서 많게는 4 프레임까지를 필요로 한다. 일반적인 수화 단어의 경우 이 정도 프레임 수로써

충분히 표현되므로 지화 단어가 전체 시스템의 성능 향상에 있어서 큰 문제점이 되고 있음을 알 수 있다. 하지만, 기존의 수화 발생기와는 달리 개발된 수화 발생기는 여러 가지 전처리 과정과 생략 기능을 통하여 지화 단어의 비율을 현저히 줄일 수가 있었으며 차후 시스템의 단어 DB 및 기타 전처리에 필요한 DB의 갱신을 통하여 많은 발전을 보일 것으로 기대된다.

전체적인 프레임 수와 수행 시간 이외에도 프레임 수(특정한 단어는 여러 개의 프레임으로 구성됨)에 따른 수행 시간 및 프레임별 처리 속도를 살펴봄으로써 시스템 성능 개선의 지표를 삼을 수가 있을 것이다. 따라서, 해당 자막 방송 프로그램에 포함된 각 단어를 그에 해당하는 프레임 별로 구분하여 필요한 프레임의 수, 표현하기 위한 시간의 분포, 각 프레임을 표현하기 위한 속도 및 수행 시간등을 비교하였다. 그림 14 (a),(b),(c),(d)는 이상의 여러 가지 실험에 대한 결과를 막대그래프 형식으로 나타낸 것이다⁷. 여기서, 0 프레임으로 표현되는 수화나 지화는 실제로는 존재하지 않으나 고유명사를 표현할 경우 상징적인 그림으로 표현하게 되므로 0 프레임으로 계산하였다.

그림 14 (a)는 프레임 별 전체 빈도수를 나타낸다. 특히 지화 단어 중에서도 2개의 음소(초성과 중성)로만 이루어진 단어의 비율이 상당히 높으며, 수화 단어의 경우 앞에서 설명한 것과 마찬가지로 대략 3~4 프레임으로 표현 가능한 경우가 많음을 알 수 있다.

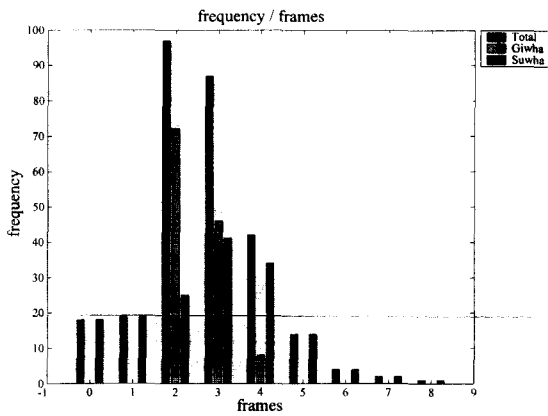
그림 14 (b)는 프레임 별 전체 수행 소요 시간을 나타낸다. 그림 14 (a)에서 알 수 있듯이 3 프레임으로 표현되는 단어들은 수화와 지화의 비율이 비슷하므로 수행 시간도 가장 오래 소요됨을 알 수 있다. 또한, 0 프레임으로 표현되는 고유명사의 경우에는 수행 소요 시간이 가장 짧은 것을 알 수 있다.

그림 14 (c)는 프레임 별 평균 수행 소요 시간을 나타낸다. 특기할 사항은 1 프레임으로 표현되는 단어의 경우, 다른 단어들에 비하여 수행 시간이 많이 걸린다는 것이다.

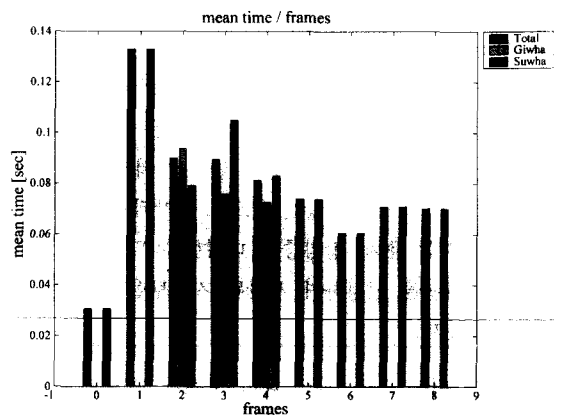
그림 14 (d)는 프레임 별 평균 수행 속도를 fps(초당 프레임)으로 나타낸 것이다. 0 프레임으로 표현되는 고유 명사의 경우 현재 수화 발생기 시스템에서 사용하고 있는 샘플링 타임인 30ms와 거의 일치하는 결과

6) 표 4.4.1에서 '전체비율'은 실제 프로그램에 포함된 단어와 수화발생기에서 생략(Skip)하지 않고 처리하는 단어와의 비율을 의미한다. 즉, '뽀뽀뽀' 프로그램에서는 33%의 단어가 생략 처리되는 것이다.

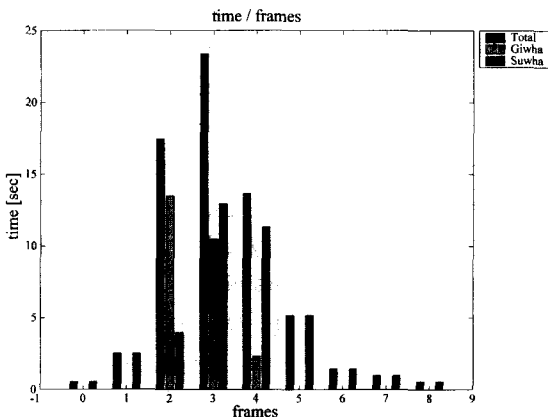
7) Total : 전체 단어(수화+지화), Giwha : 지화, Suwha : 수화를 각각 의미한다.



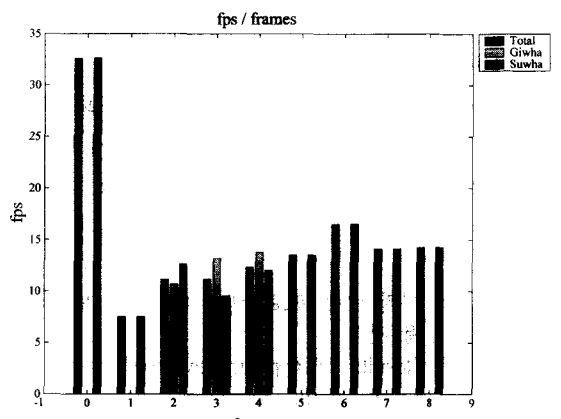
(a) 프레임별 전체 빈도수



(c) 프레임별 평균 수행 소요 시간



(b) 프레임별 전체 수행 소요 시간



(d) 프레임별 평균 수행 소요 속도

그림 14. 제안된 시스템의 성능에 대한 통계치 분석

Fig 14. Statistical analysis of performance of the proposed system.

(1/30ms = 약 33 fps)를 보이고 있다. 그 이외의 경우에도 1프레임으로 표현되는 단어들을 제외하고는 11 fps에 근접한 결과를 보여 주고 있다.

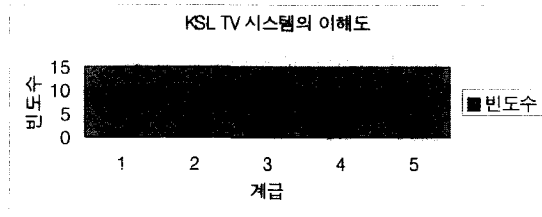
2. 농아인들의 평가

개발된 시스템은 농아인들을 위한 다수의 행사에 전시되어 농아인들이 직접 시스템을 사용한 후 다양한 평가 및 의견을 제시하였다. 시스템에 대한 평가는 설문조사의 형식을 취하였으며, 주요 평가 항목은 자막 방송에 대한 이해도, 시스템 화면 구성 및 동작에 대한 친숙도, 시스템 동작 속도로 선택하였다(그림 15 (a),(b),(c),(d))⁸.

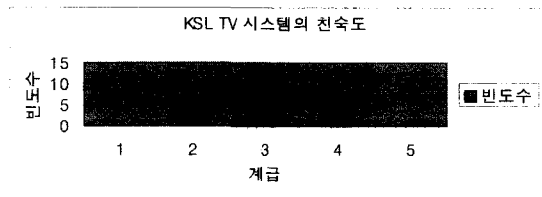
이해도 측면에서는 대다수의 농아인들이 기존의 자막 방송 프로그램 혹은 수화 통역사를 동반한 방송 프로그램에 비해서 수화 내용을 이해하기가 비교적 용이한 것으로 답변하였으며(그림 15 (a)), 친숙도 측면에서도 예상했던 바와 같이 인간 친화적인 것으로 파악되었다(그림 15 (b)). 한글 수화 발생기의 사용자 인터페이스 구성(화면 구성)의 경우 역시, 전반적으로 잘 구성되었다고 평가하였다(그림 15 (c)). 다만, 시스템의 동작 속도에 있어서는 기존의 TV 방송에 비하여 조금 빨라서 적응이 힘들다는 의견도 제시되었다(그림 15 (d)). 이것은 3차원 애니메이션 방식의 개선과 3차원 모델의 교환 등으로 쉽게 향상 가능한 부분으로 생각된다.

8) 그림 15에서 '계급'은 각 항목에 대한 평가 점수를 나타내며, '빈도수'는 각 계급에 대하여 응답한 사람의 수를 나타낸다.

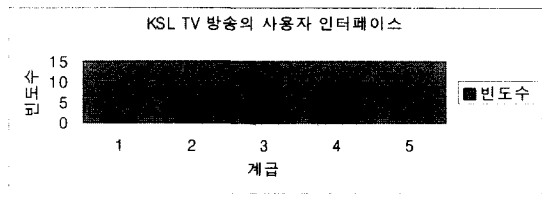
그 외의 의견으로는 좀 더 인간 친화적인 3차원 캐릭터를 구성하기 위한 얼굴 표정의 다양화, 실제 사람과 비슷한 캐릭터의 사용 등이 언급되었다.



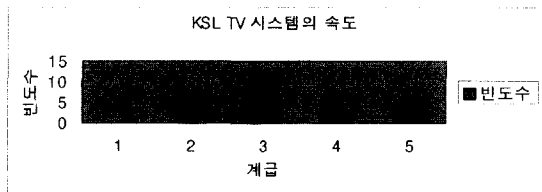
(a)



(b)



(c)



(d)

그림 15. 실제 농아인들을 대상으로 한 개발된 시스템에 대한 평가 결과.

Fig. 15. Evaluation results of the developed system with the real hearing-impaired.

V. 결론 및 추후과제

본 논문에서는 자막 방송을 위한 한글 수화 발생기의 전체 구성에 대하여 세부적으로 설명하고, 자막 방송 프로그램에 대한 적용을 통해 그 유용성을 보였다. 실제 농아인들을 대상으로 한 시연을 통하여 개발된

시스템에 대한 긍정적인 반응을 얻을 수 있었다. 제안된 형태소 분석기는 수화의 언어적 특성에 기반한 각종 전처리 방법을 효과적으로 사용함으로써, 전체 시스템의 속도 향상과 이해도 측면에서 많은 기여를 하고 있음을 알 수 있었다. 또한, 개발된 한글 수화 발생기는 자막 신호에 따른 표정 변화, 캐릭터 얼굴의 변화, 고유 명사의 표현 등을 통하여 보다 인간 친화적인 시스템을 구현할 수 있었다.

개발된 시스템의 시연을 통해 수렴된 농아인들과 농아인 교육에 종사하는 수화 전문가들의 의견들은 크게 다음과 같이 요약될 수 있었다.

1) 개발된 시스템을 적용할 경우, 때에 따라서는 현재와 같이 제한된 형태의 자막 방송이 아닌 24시간 무제한으로 방영되는 자막 방송 프로그램의 구성이 가능해 질 것이다.

2) 보다 자연스러운 의미 전달을 위해서는 3차원 수화 캐릭터와 자막 방송 화면간의 동기화(Synchronization) 문제가 중요한 것으로 판단된다.

3) 자막 방송과 한글 수화 발생기의 결합에 의한 파급 효과는 예상 외로 클 것이다. 실제로 국내에 거주하는 농아인들 중에서 교육의 혜택을 입은 4분의 1 정도에 해당하는 농아인들만이 우리 글을 읽고 의미를 파악할 수 있다. 하지만, 이들도 현재의 자막 방송 프로그램과 같이 자막 정보가 급하게 표시되는 경우 많은 어려움을 느끼고 있다. 이러한 맥락에서 개발된 시스템은 농아인들로 하여금 보기 편하고 쉬운 방송 프로그램을 구성한다는 측면에서 매우 유용할 것으로 판단된다.

4) 또한, 학교 교육을 제대로 받지 못한 일반적인 농아인들이나 중복 장애인들의 경우 개발된 시스템을 사용하면 프로그램에 대한 이해도 증가 이외에도, 우리말 습득에 있어서 많은 교육 효과를 가져올 것으로 기대된다. 이것은 비단 교육의 기회 증대 뿐만 아니라 폭넓은 혹은 원만한 사회 생활을 영위하기 위한 필수적인 수단으로 생각된다.

제안된 자막 방송 기반 한글 수화 발생기의 상용화 및 성능 개선을 위해서는 수화 DB 및 전처리 관련 DB의 확장, 수화적 특성에 기반한 자막 방송 표준안 마련 등이 필요할 것으로 생각되며, 이를 위해서는 방송 제작 관련 업체와의 협의가 필요할 것으로 생각된

9) 장애를 두 가지 이상 가지고 있는 중증 장애인 등

다. 구체적으로 수화에 적합한 단어를 사용한 자막 방송 대본의 제작, 가능한한 쉬운 우리말 문법의 적용, 농아인들의 독해력 수준을 고려한 자막 방송 대본의 순화 작업 등은 반드시 필요한 작업이 될 것이다. 또한, 장기적으로는 문장 기반의 형태소 분석을 통한 상황 기반(Context-based) 형태소 분석기법이라든가, 동음이의어(Homonym)의 처리를 통한 형태소 분석 기능의 향상 등이 이루어져야 할 것이다.

참 고 문 헌

[1] 이은표, “자막 방송 표준에 대한 소개”, 정보통신 표준화 위원회 활동소식, 1998
 [2] 한국정보통신기술협회, “텔레비전 자막방송 잠정 표준”, TTA.KO-07.0010, 1997

[3] 안일남, “정보화와 청각장애인”, 정보와 청각장애인, 1997
 [4] 변중남 외, “청각 장애인 교육을 위한 수화 통역 시스템의 개발”, 삼성전자주식회사, 2000
 [5] <http://csfive.kaist.ac.kr/>, 한글형태소분석기, 한국과학기술원 최기선 교수 연구실, 2000
 [6] 석동일, “한국수화의 언어학적 분석”, 대구대학교 박사학위논문, 1989
 [7] 울산대학교, 한국과학기술원, “자연언어/블리언 질의 변환 시스템”, (주)테이콤, 1990.
 [8] <http://kordic.britannica.co.kr/>, 2000.
 [9] 변중남 외, “지능적 인식 및 제어에 관한 연구 III”, 삼성전자주식회사, 1996
 [10] 변중남 외, “지능적 인식 및 제어에 관한 연구 II”, 삼성전자주식회사, 1994

저 자 소 개



金大珍(正會員)
 1997년 2월 경북대학교 전자공학과 (학사), 1999년 2월 한국과학기술원 전기및전자공학과 (석사), 1999년 3월~현재 : 한국과학기술원 전자전산학과 전기및전자공학전공 박사과정, <주관심분야 : 영상처리, 얼굴인식 및 표정인식, 인간-로봇 상호작용 기술, 소프트 컴퓨팅>

金亭培(正會員) 第36卷 第6號 參照

卞增男(正會員) 第36卷 第6號 參照

張源(正會員) 第35卷 C編 第6號 參照