

아나운서와 음성합성기의 뉴스 낭독 음성 비교 연구 -유사언어에서의 차이를 중심으로-

유지철(성균관대학교)

Comparisons of Paralanguage Features between a Human Announcer and Text-to-Speech(TTS) Devices during the Out-Loud Reading of News Sentences. Yu, Jeechul. <Abstract> I studied the paralanguage differences between a human announcer and TTS devices during the out-loud reading of news sentences, and arrived at the following results. First, the announcer's intonation changed flexibly within a sentence whereas the intonation of the TTS devices was rather fixed at an either rising or falling tone. Second, the announcer stressed specific words within a sentence by changing her voice intensity, whereas the TTS devices maintained the same intensity throughout the sentence. Third, pause duration differed for both of the announcer and the "nVoice"-type TTS device but not for the "VoiceText"-type TTS device. In addition, the TTS devices made errors in reading by punctuating. Fourth, the announcer kept changing the pace within a sentence, whereas the TTS devices read all the words at an almost constant pace. Fifth, the announcer pronounced short and long vowels distinctively, but the TTS devices pronounced all vowels almost always using the same length. Sixth, the announcer widened her mouth when pronouncing open vowels, corresponding with the opening widths of given vowels, whereas the opening width of the TTS devices was smaller compared to that of the announcer's. In sum, when reading aloud news sentences, the human announcer controlled the paralanguage features in a context-dependent manner, whereas the TTS devices tended to maintain the same paralanguage features throughout entire sentences.

Key Words paralanguage(유사언어), intonation(억양), intensity(강도), pause(휴지), pace(전달 속도), length(음장), opening width(개구도)

I. 서론

이 연구의 목적은 음성합성기와 아나운서가 동일한 뉴스를 낭독한 음성의 유사언어를 비교, 분석함으로써 그 분석 결과를 바탕으로 음성합성기 합성음이 아나운서의 목소리와 같은 자연성과 명료도를 실현하기 위해서는 유사언어를 어떻게 조절하여야 하는지에 대한 시사점을 제시하는 데 있다.

이 연구에서는 음성합성기가 사람과 같이 유사언어를 조절하면서 뉴스를 낭독하는지를 고찰하기 위하여 음성합성기의 합성음과 실제 사람의 뉴스 낭독 음성을 음성 분석 프로그램인 Praat를 이용하여 실증적으로 비교, 분석하고자 한다. 이 연구에서는 사람의 표본으로 아나운서를 선정하였다. 아나운서는 전문적으로 뉴스를 전달하는 사람으로서 평소에 언어와 비언어에 대한 훈련을 받고 실제 방송을 통해 풍부한 뉴스 낭독 경험을 갖추고 있어 언어적, 비언어적 메시지를 가장 효과적으로 전달할 수 있는 사람으로 평가를 받고 있기 때문이다.

음성합성기는 텍스트 문장을 음성으로 변환시켜주는 음성 합성 기술인 TTS(Text-to-speech)를 구현하는 소프트웨어 및 하드웨어를 말한다. 음성 합성 기술은 기술의 완성도가 날로 높아져 금융, 통신, 유통 등 서비스 산업에서부터 컴퓨터, 가전제품, 스마트폰에 이르기까지 음성 출력 수단으로 다양하게 활용되고 있다(이종석, 박기태, 이준우, 2014, p. 204). 일례로 인터넷 포털 사업자인 네이버사는 포털을 통해 제공하는 모든 뉴스 텍스트에 대해 ‘본문듣기’라는 오디오 지원 서비스를 실시하여 이용자들이 편리하게 뉴스

텍스트를 음성을 통해 들을 수 있게 되었다. 그런데 음성합성기의 음성은 실제 녹음된 사람의 목소리가 아니라 TTS 기술을 사용해 합성한 목소리이기 때문에 실제 사람의 음성과는 차이가 있다. 양병곤(2002, p. 90)에서는 사람처럼 말하는 음성 합성 컴퓨터가 개발되고 있지만 이런 컴퓨터가 사람처럼 자연스럽게 말하기에는 아직 역부족이라고 하면서 그만큼 사람의 목소리는 복잡하고 정교한 반면 음성 합성 기술은 여전히 사람의 발음기관의 기능을 제대로 반영하지 못한다고 한다.

일반적으로 음성합성기의 합성음은 사람의 목소리 운율과 얼마나 비슷한지를 평가하는 자연성(naturalness)의 측면과 합성음의 발음이 얼마나 정확하게 정보를 전달하는지를 평가하는 명료도(intelligibility)의 측면에서 평가될 수 있다(조철우, 이상호, 김수진, 2005, p. 102).

뉴스 문장 낭독은 오로지 목소리를 통해서만 정보를 전달하기 때문에 목소리의 구성 요소인 유사언어(paralanguage)¹⁾를 어떻게 조절하여 낭독하느냐가 매우 중요하다. 같은 내용의 뉴스라고 하더라도 화자가 뉴스를 어떻게 낭독하느냐에 따라 이를 청취하는 수용자의 뉴스 이해도와 호감도에 큰 차이가 발생할 수 있다. 뉴스 문장을 의미에 맞게 끊어 읽고 끊어 읽는 중간에 적당한 휴지(pause)를 주며 문장의 내용에 따라 낭독 속도의 완급을 조절하고 핵심이 되는 단어와 구를 강도를 높여 낭독하며 뉴스의 분위기에 맞게 억양을 변화시키는 등 유사언어를 조절하면서 뉴스 문장을 낭독하여야 수용자가 뉴스의 내용을 잘 이해할 수 있고 호감도도 높아질 수 있다. 그렇지 않고 뉴스의 내용과 의미를 제대로 이해하지 않은 채 유사언어에 변화를 주지 않고 기계적으로 단조롭게 뉴

1) 유사언어는 음성언어 중 어휘, 발화, 메시지 구성과 같은 언어적 요소를 제외한 발음, 말의 속도, 목소리의 크기, 어조, 억양, 강세 등 언어적 요소와 불가분의 관계에 있는 요소로서 준언어, 부차언어라고도 불리며 우리가 발화하는 순간 언어적 요소와 동시에 실현된다(이도영, 2002, pp. 25-45). 이 연구에서는 이를 ‘유사언어’로 지칭하기로 한다.

스 문장을 읽을 경우 이는 단순한 글자 읽기에만 그칠 뿐 청자에게 뉴스를 전달하는 데에는 실패할 수밖에 없다. 그리하여 이 연구에서 음성합성기의 합성음이 아나운서의 목소리와 같은 자연성과 명료도를 실현하기 위해서는 유사언어를 어떻게 조절하여야 하는지에 관해 고찰하고자 한다.

II. 선행 연구 고찰

선행 연구를 유사언어의 구성 요소들 중에서 Praat로 실증적인 분석이 가능한 요소들인 억양, 강도, 휴지, 전달 속도, 음장, 발음의 개구도 등을 중심으로 살펴보하고자 한다.

1. 억양

문장 낭독과 관련하여 정동빈(1997, p. 38)에서는 문장의 구나 절의 끝 억양은 끝이 올라가는 상승조와 평탄하게 이어지는 평탄조, 아래로 내려가는 하강조가 있는데, 낭독을 처음 할 때에는 끝 억양을 모두 올리는 경향이 있어 초보 단계에서는 가능하면 내리도록 노력해야 하고, 뉴스 방송이 익숙해지면 상승조와 평탄조, 하강조의 변화를 적절히 조절하여야 한다고 한다.

남도현, 임성수, 윤보람, 최홍식, 유애리(2010, p. 96)에서는 일반인 화자와 아나운서의 문장 읽기에서의 기본주파수 범위에 대한 실험 연구에서 일반인 남성은 87.7Hz-153Hz의 범위에서 낭독을 하는 데 비해 남성 아나운서는 87.3Hz-180.2Hz의 범위에서 낭독을 하고, 일반인 여성은 157.8Hz-286.8Hz의 범위에서 낭독을 하는 데 비해 여성 아나운서는 148.2Hz-331.2Hz의 범위에서 낭독을 한다는 실험 결과를 통하여 남녀 아나운서들은 일반인 남성과 여성에 비해 문장 읽기에서 훨씬 넓은 기본주파수 범위에서 음높이의 변화를 보이며 문장을 읽는다고 한다.

2. 강도

전은주(1991, pp. 45-47)에서는 대학생 화자들을 대상으로 한 한국어 초점(focus)에 대한 실험음성학적 연구를 통하여 문장 내에 어느 단어가 초점을 이룰 때 초점이 아닌 단어들과 비교해 음향음성학적으로 강도(intensity)가 증가한다는 사실을 밝혀내었는데, 초점을 받는 명사는 58.6dB의 강도로 발화하는 데 비해 중립 명사는 57dB로 발화하고 초점을 받지 않는 명사는 56.3dB의 강도로 발화한다고 한다.

남도현 외(2010, pp. 84-88)에서는 일반인 화자와 아나운서의 문장 읽기에서의 소리의 강도 범위에 대한 실험 연구에서 일반인 남성은 52dB-63dB의 범위에서 낭독을 하는 데 비해 남성 아나운서는 56dB-70.9dB의 범위에서 낭독을 하고, 일반인 여성은 49.6dB-59.2dB의 범위에서 낭독을 하는 데 비해 여성 아나운서는 54.7dB-66dB의 범위에서 낭독을 한다는 실험 결과를 통하여 남녀 아나운서들이 일반인 남성과 여성에 비해 문장 읽기에서 훨씬 넓은 강도 범위에서 강도의 변화를 보이며 문장을 낭독한다고 한다.

3. 휴지

이호영(1991, pp. 3-43)에서는 올바른 숨쉬기와 끊어 읽기는 청자들에게 자연스럽고 편안하게 내용을 전달하기 위한 필수적인 요소라고 하면서 숨쉬기와 끊어 읽기를 하는 위치는 문장의 문법 구조, 말의 속도, 스타일, 초점(focus)의 영역 등에 따라 결정된다고 한다.

지민제, 한성숙, 김윤기(1996, p. 69)에서는 말의 경계와 휴지의 길이 사이의 상관관계를 밝혀내었는데, 뉴스를 비롯한 실제 발화를 분석한 결과 약 25%가 휴지 시간으로 나타났으며 단어 경계와 구 경계 사이에는 휴지의 길이 차이가 거의 없으나 구, 절, 문장, 단락의 상위 경계로 올라갈수록 휴지의 길이가 길어짐을 확인하였다.

4. 전달 속도

박경희(2004, pp. 69-71)에서는 뉴스 전달 속도에 따른 청자들의 청취 소감을 조사하여 1분당 300음절 이하가 되면 활력이 없고 느리게 감지되고, 350음절 정도이면 차분하다고 인식되어 뉴스 이해도가 높으며, 400음절 내외에서는 빠르다는 느낌과 함께 음가가 탈락, 왜곡되는 현상이 나타나면서 내용 파악에 어려움을 겪는다고 한다.

전은주(1991, pp. 21-35)에서는 문장 내에 어느 단어가 초점을 이룰 때 초점이 아닌 단어들과 비교해 음향음성학적으로 길이, 기본주파수, 강도가 증가한다고 한다. 이 3가지 요소 중 가장 일반적으로 변화를 보이는 것은 길이라고 한다. 화자들은 초점이 되는 단어를 그렇지 않은 단어보다 발화 속도를 느리게 하여 길이를 더 길게 낭독함으로써 초점이 되는 단어가 돋들리게 한다고 한다. 실험결과 초점을 받는 단어는 565msec의 길이로 발화하는 데 비해 중립 단어는 484msec로 발화하고 초점을 받지 않는 단어는 489msec로 발화한다고 한다.

5. 음장

지민제 외(1996, pp. 65-66)에서는 ‘말[言]’과 ‘말[馬]’ 등 같은 음소를 사용하지만 앞의 것은 장모음 /ㅏ:/를 쓰고 뒤의 것은 단모음 /ㅏ/를 써서 뜻을 구별하는 것처럼 길고 짧은 소리 길이의 차이로 뜻이 갈릴 때, 이를 ‘장단 대립(長短對立)’이라고 한다고 정의한다.

박경희(2003, pp. 32-54)에서는 같은 30대 연령대의 아나운서 그룹과 일반인 그룹의 장단 모음 간 음장 상댓값을 조사한 결과 아나운서 그룹은 장단 모음 간 상댓값이 ‘ㅏ:/ㅏ(1.55)’, ‘ㅓ:/ㅓ(1.40)’, ‘ㅕ:/ㅕ(1.33)’, ‘ㅗ:/ㅗ(1.94)’로 변별력이 높지만 일반인 그룹은 장단 모음 간 상댓값이 ‘ㅏ:/ㅏ(1.09)’, ‘ㅓ:/ㅓ(0.90)’, ‘ㅕ:/ㅕ(0.64)’, ‘ㅗ:/ㅗ

(1.30)’으로 변별력이 낮거나 오히려 단모음을 장모음보다 길게 발음하는 것을 확인하였다.

6. 발음의 개구도

양병곤(2002, p. 91)에서는 사람은 혀나 턱, 입술과 같은 조음기관을 끊임없이 움직이면서 발음을 한다고 하면서 모음을 발음할 때 입 안의 모양을 보면 ‘아’를 발음할 때에는 입을 많이 벌리고 혀가 낮게 깔리고 ‘이’를 발음할 때에는 입을 거의 다물고 혀의 앞부분을 입천장 쪽으로 가까이 둔다고 한다. 이처럼 모음의 발음은 발음을 할 때 혀의 최고점의 높이가 입천장으로부터 얼마나 떨어져 있는가, 즉 턱을 얼마나 벌리는가에 해당하는 개구도(開口度)에 의하여 결정되는데, ‘ㅏ’, ‘ㅑ’, ‘ㅕ’와 같이 혀의 최고점이 낮은 저모음은 턱을 벌리고 발음해야 정확한 모음의 음가를 실현할 수 있다.

이상의 선행 연구들을 종합해 보면 아나운서는 일반인 화자들에 비해 유사언어 요소들을 더욱 폭넓게 변화시키며 문장을 낭독한다는 사실을 실험적인 연구를 통하여 증명하였다. 그러나 아직 음성합성기의 합성음과 사람의 음성에 대한 비교 연구는 치밀하게 이루어져 있지 않다. 이 연구에서는 아나운서와 음성합성기의 뉴스 낭독 음성을 음성 분석 프로그램을 이용하여 분석함으로써 유사언어의 차이를 실증적으로 밝혀내고자 한다.

Ⅲ. 연구 방법

1. 실험 뉴스 녹음물 제작

아나운서와 음성합성기가 동일한 내용의 뉴스 문장을 낭독할 때 유사언어에서 어떤 차이가 있는지를 살펴보기 위하여 아나운서와 음성합성기의 뉴스 낭독 녹음물 제작이 필요하였다. 녹음에 사용

한 실험 뉴스 문장은 아래와 같다.

미국의 금리 인상과 관련해 세계 금융시장이 흔들린 지난달, 외국인 주식 투자 자금이 4년여 만에 가장 많이 빠져나간 것으로 나타났습니다. 한국은행이 지난 8일 발표한, 2018년 2월 중 국제 금융 외환 시장 동향을 보면 지난달 국내에서 빠져나간 외국인 주식자금은 36억 3천만 달러로, 2013년 6월 이후 최대 규모를 기록했습니다. 2013년 6월은 미국 연방준비제도가 양적 완화 정책을 긴축으로 전환하면서 국제 금융 시장이 크게 흔들렸던 때입니다. 한국은행 관계자는 당시와 달리 지금의 외국인 자금 유출은 위험 회피 성향 강화에 따른 단기적인 성격이 강하다며 2월 중순 이후 불안 심리가 회복되면서 증권 투자 자금 유입이 우세했다고 설명했습니다. 세계 외환 시장 변동성이 커지면서 지난달 말, 원 달러 환율은 1월보다 14원 90전 오른, 달러 당 1,082원 80전을 기록했습니다. 전일 대비 원 달러 환율 변동 폭은 5원 50전으로, 1월의 3원 80전보다 커졌습니다.

먼저 KBS 9시 뉴스를 진행한 경력이 있는 KBS 여성 아나운서가 녹음실에서 뉴스를 녹음하였고, 같은 내용의 뉴스 문장을 보이스웨어사에서 개발한 음성 합성 엔진인 VoiceText의 여자 버전인 Gyuri와 Yumi 버전과 네이버사에서 개발한 음성 합성 엔진인 nVoice의 여자 버전으로 녹음하여 총 4개의 뉴스 녹음물을 제작하였다. 음성 분석의 정확성을 높이기 위하여 4개의 녹음물을 대상으로 음성의 크기인 음량을 동일하게 조정하는 작업을 거쳐 최종적으로 뉴스 녹음물을 WAVE 파일로 제작하였다. 실험 대상 화자는 다음 Table 1.과 같다.

Table 1. experimental speakers

Speaker	Class	Affiliation	Sex
Announcer	Human	KBS	Female
Gyuri	TTS device(VoiceText)	Voiceware	Female
Yumi	TTS device(VoiceText)	Voiceware	Female
nVoice	TTS device(nVoice)	Naver	Female

2. 음성 분석 방법

이 연구에서는 음성 분석 프로그램인 Praat²⁾(ver.6.0.3.9)를 이용하여 뉴스 녹음물의 억양, 강도, 휴지, 전달 속도, 음장, 발음의 개구도 등을 비교, 분석하였다.

4개의 뉴스 녹음물의 발화 내용을 각각 어절이나 음절 단위로 분절(segmentation)하고 주석을 표기(labelling)한 다음 피치 값 분석을 통하여 억양을 비교하고 강도 값 분석을 통하여 강도를 비교하였으며 음성 파형의 묵음 구간 시간을 측정하여 휴지 시간을 비교하였다. 또한 음성파형의 조음 음절 수와 조음 시간을 측정하여 전달 속도를 비교하였고 음성파형과 스펙트로그램(spectrogram) 분석을 통하여 음장을 비교하였으며 끝으로 포먼트(formant) 분석을 통하여 발음의 개구도를 비교하였다.

억양과 강도를 비교하기 위해서는 음성의 피치 값 및 강도 값의 변동 범위를 수치화하여 비교하여야 하는데, 이를 위하여 각 음성의 피치 값과 강도 값 리스트를 마이크로소프트사의 엑셀(Excel) 프로그램의 box-and-whisker plot에 입력하여 피치 값 및 강도 값의 변동 범위를 통계 그래프로 나타내었다.

IV. 음성 분석 결과 및 논의

1. 억양

억양은 음높이, 즉 피치(pitch)의 변화이므로 이 연구에서는 4개

- 2) Praat는 말소리의 음성 과학적 분석을 위한 컴퓨터용 공개 소프트웨어이다. 네덜란드 암스테르담 대학교의 파울 부르스마(Paul Boersma)와 다비트 베닝크(David Weenink)가 개발한 프로그램으로 여러 운영체제에서 실행 가능한 음성 분석 프로그램이다. 음성 파형을 스펙트로그램(spectrogram), 피치(pitch), 강도(intensity), 포먼트(formant), 펄스(pulses) 등으로 분석할 수 있다(조동욱, 2009, p. 52).

의 음성 자료에 대한 피치 분석을 통해 억양의 변화를 살펴보았다. 분석 결과 아래 Figure 1.과 Table 2.와 같이 아나운서의 피치 값 전체 변동 폭이 가장 크게 나타났고 이어서 nVoice, 규리, 유미의 순이었다. 상자그림으로 표시된 1사분위에서 3사분위까지의 피치 변동 폭인 사분편차(quartile deviation)는 아나운서와 nVoice가 가장 크게 나타났고 이어서 규리, 유미의 순이었다.

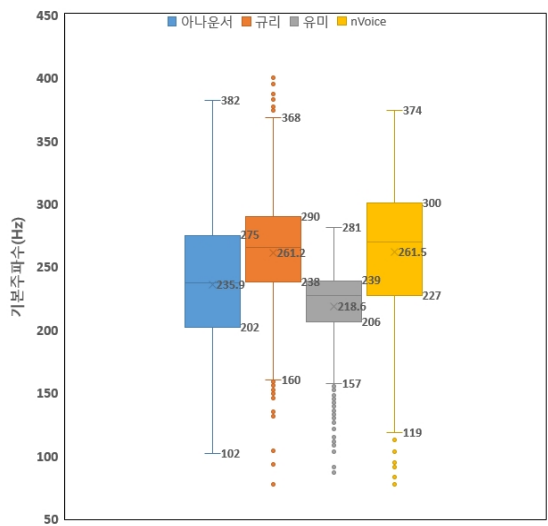


Figure 1. variability of pitch across-speakers

Table 2. average and variability of pitch across-speakers(unit: Hz)

Speaker	Average pitch	Variability of pitch (minimum-maximum)	Variability of pitch (1quartile-3quartile)
Announcer	236	280(102-382)	73(202-275)
Gyuri	261	208(160-368)	52(238-290)
Yumi	219	124(157-281)	33(206-239)
nVoice	262	255(119-374)	73(227-300)

각 화자별 피치 곡선을 통해 나타난 특징을 보면 아나운서는 아래 Figure 2.처럼 문장 내 각 구의 처음에 오는 성분은 높은 피치의 억양으로 시작하는 반면에 구가 끝나는 성분에서는 하강조의 억양을 구사하는 등 억양에 변화를 주었다. 즉 각 구의 처음에 오는 성분을 낭독할 때는 새로운 정보로 취급하여 높은 피치를 부여해 강조함으로써 청자의 주의를 끌었고 구가 끝날 때는 하강조로 억양을 구사하면서 마무리하였다. 박경희(2004, pp. 29-30)에서 억양은 화자의 태도를 드러내 상승 억양은 미완의 뜻을 나타내고 하강 억양은 완결의 의미를 나타낸다고 하였는데, 아나운서는 문장 내의 각 구를 시작할 때에는 상승조로 낭독하여 미완의 뜻을 나타내고 구가 끝날 때는 하강조로 낭독하여 완결의 의미를 나타내는 것으로 해석할 수 있다.

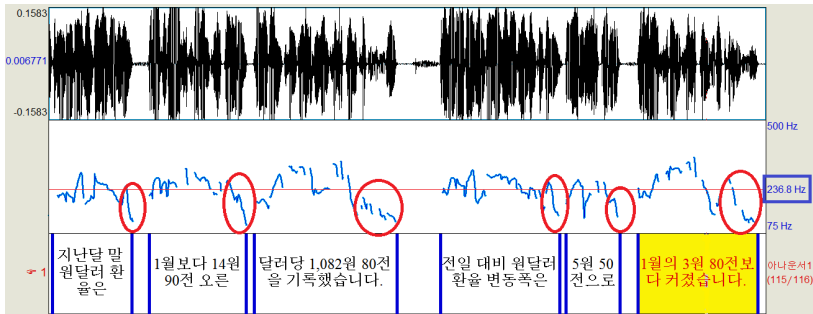


Figure 2. the pitch curve of the announcer

규리는 Figure 3.처럼 아나운서와 같이 비교적 폭넓게 음높이를 변화시키며 뉴스를 낭독하였다. 그러나 아나운서가 상승조와 하강조의 억양 변화를 주는 것과 달리 규리는 문장 내의 각 구가 시작할 때와 구가 끝날 때 모두 상승조의 억양을 구사하였다.

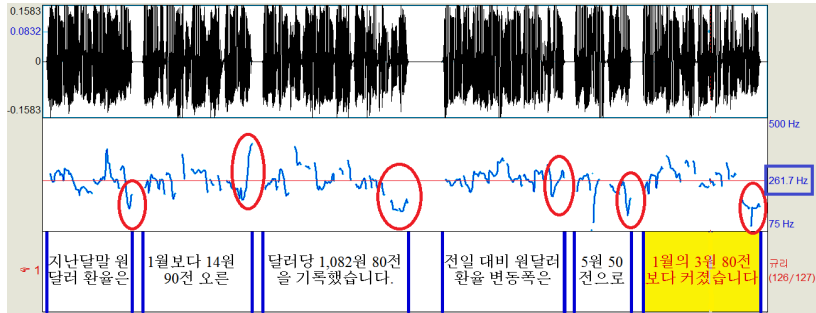


Figure 3. the pitch curve of Gyuri

유미는 피치 변동 폭이 가장 작았는데, 아래 Figure 4.에서처럼 음높이를 많이 변화시키지 않고 문장을 낭독하였다. 또한 아나운서는 구가 시작할 때는 상승조의 억양을 구사하고 구가 끝날 때는 하강조의 억양을 구사하는 등 억양의 변화를 준 데 비해 유미는 전체적으로 음높이의 변화가 없이 평탄조로 낭독하다가 구가 끝날 때 하강조로 마무리하는 단조로운 낭독을 하였다.

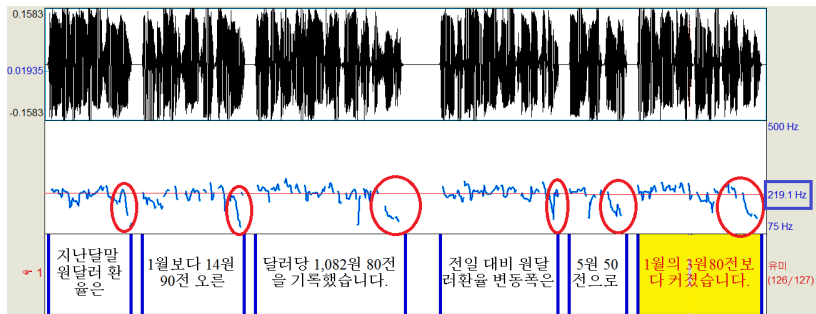


Figure 4. the pitch curve of Yumi

nVoice는 아래 Figure 5.에서처럼 아나운서와 같이 폭넓게 음높이를 변화시키며 뉴스를 낭독하였다. 그러나 아나운서가 상승조와 하강조의 억양 변화를 주는 것과 달리 nVoice는 규리와 같이 문장

내의 각 구가 시작할 때와 구가 끝날 때 모두 상승조의 억양을 구사하였다. 특히 nVoice는 각 구가 끝나는 성분을 낭독할 때 급격한 상승조 억양을 구사하는 특징을 보였다.

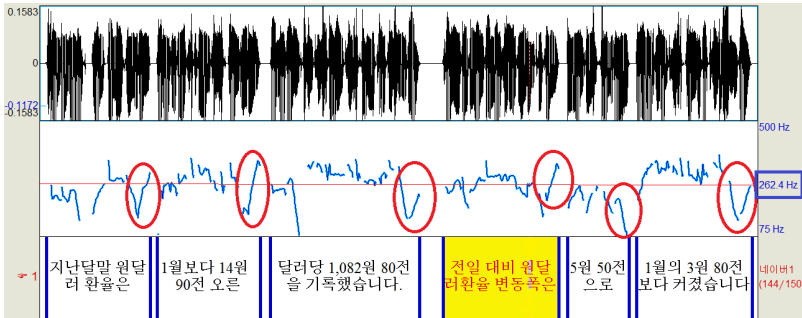


Figure 5. the pitch curve of nVoice

정리하면 아나운서와 nVoice, 규리는 비교적 폭넓게 음높이를 변화시키며 뉴스를 낭독하는 반면 유미는 상대적으로 음높이 변화가 적게 문장을 낭독하였고 아나운서는 구가 시작할 때는 상승조의 억양, 구가 끝날 때는 하강조의 억양을 구사하는 등 억양의 변화를 준 데 비해 nVoice, 규리는 전체적으로 상승조 억양을, 유미는 전체적으로 평탄조 및 하강조 억양을 구사하는 특징을 보였다. 이처럼 음성합성기의 피치 변동 폭에 차이가 발생하고 음성합성기마다 상승조 혹은 하강조의 억양 차이가 발생하는 이유는 합성음에 성우들의 억양이 그대로 실려 있기 때문이다.

음성 합성 기술은 전문 성우의 목소리로 음편들을 사전에 녹음해 음성 DB에 저장해 두고 텍스트가 입력되면 텍스트와 가장 유사한 음편들을 불러와 연결해 편집하는 형식인데, 음편들을 연결하는 과정에서 억양과 강세, 휴지의 길이, 음소의 지속 시간 등의 운율을 별도로 처리하지 않고 이미 운율이 실려 있는 음편 중에 최적의 운율을 포함하는 음편을 선택하여 합성하는 방식이다(이중

석 외, 2014, p. 30). 따라서 음성 DB의 음편을 녹음하는 성우가 끝
억양을 상승조나 하강조로 낭독할 경우 억양이 음편에 남아있게
되어 음성합성기가 음성 DB에서 음편을 불러와 합성하면 음편에
남아 있던 상승조나 하강조의 억양이 그대로 나타나게 된다.

2. 강도

음성과형에서 강도(intensity) 값은 소리의 크기를 나타내는 척도
로서 데시벨(dB)을 단위로 한다. 강도 값은 음성과형의 진폭 값으
로 구하는데, 진폭 값이 크면 소리가 세지고 진폭 값이 작으면 소
리가 약해진다(양병곤, 2010, p. 101).

각 화자별 강도 값 변동 폭을 살펴보면 아래 Figure 6.과 Table
3.과 같이 아나운서의 강도 값 변동 폭이 전체 변동 폭과 사분편차
모두 가장 크게 나타났고 이어서 nVoice, 규리, 유미의 순이었다.

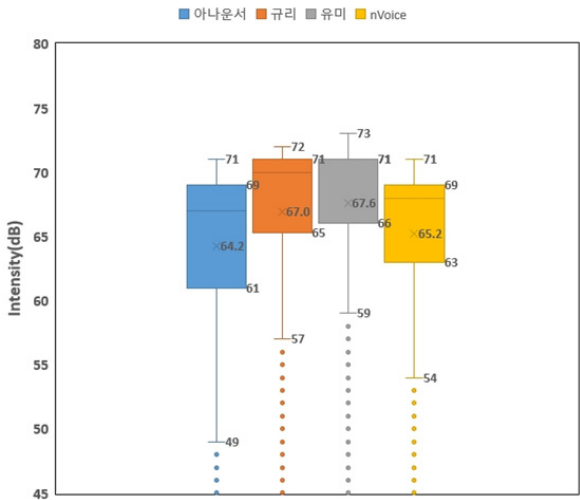


Figure 6. variability of intensity across-speakers

Table 3. average and variability of intensity across-speakers(unit: dB)

Speaker	Average intensity	Variability of intensity (minimum-maximum)	Variability of intensity (1quartile-3quartile)
Announcer	64.2	22(49-71)	8(61-69)
Gyuri	67.0	15(57-72)	6(65-71)
Yumi	67.6	14(59-73)	5(66-71)
nVoice	65.2	17(54-71)	6(63-69)

화자별 강도 곡선을 통해 나타난 특징을 보면 아나운서는 아래 Figure 7.과 같이 정보를 가지고 있는 단어들은 일정한 강도를 유지하며 낭독한 반면, 정보를 가지고 있지 않은 단어들은 강도를 줄여 낭독하였는데, 정보를 가지고 있지 않은 마지막 서술어인 ‘나타났습니다’를 낭독할 때 음성과형의 진폭이 줄어들고 강도 곡선이 낮아지는 것을 확인할 수 있다.

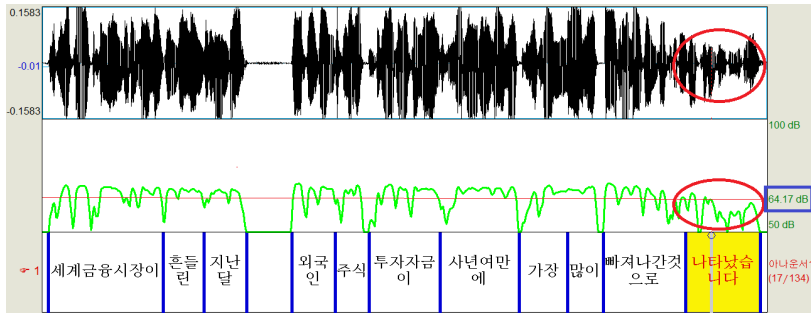


Figure 7. the intensity curve of the announcer

한편 규리와 유미, nVoice는 아나운서와 비교해 음의 강도를 많이 변화시키지 않고 뉴스를 낭독하였다. 강도 곡선을 통해 나타난 특징을 보면 음성합성기는 아나운서와 달리 문장 내의 모든 단어들을 일정한 강도로 낭독하였다. 아래 Figure 8.을 보면 규리는 아나운서가 강도를 줄여 낭독한 ‘나타났습니다’를 낭독할 때 음성과형의 진폭이나 강도 곡선이 줄어들지 않고 일정하게 유지되는 것을 확인할 수 있다.

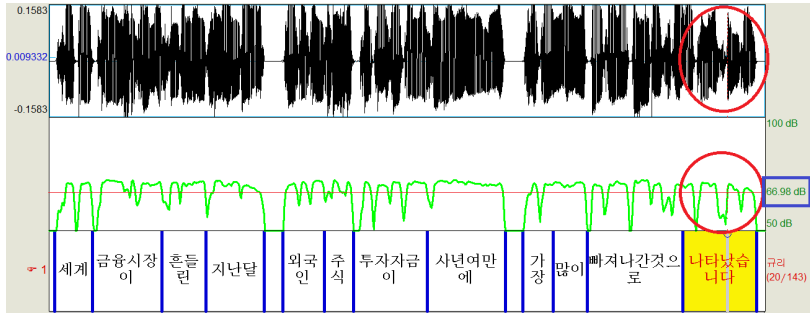


Figure 8. the intensity curve of Gyuri

이처럼 음성합성기 합성음의 음의 강도가 큰 변화 없이 일정한 모습을 보이는 이유는 음성 DB에 저장할 음편들을 녹음할 때 성우들이 음의 강도가 일정하게 유지되도록 일관성을 유지하여 녹음하였고 음성 DB에서 음편들을 불러와 연결하는 과정에서 강세 등의 운율을 별도로 처리하지 않고 이미 운율이 실려 있는 음편을 선택하여 합성하기 때문이다(이종석 외, 2014, pp. 30-31).

3. 휴지

음성파형에서 휴지는 묵음 구간으로 표시된다. 휴지 시간은 음성파형에서 발화가 끝나고 묵음 구간이 시작되는 지점에서부터 다시 발화가 시작되는 지점까지의 길이로 측정할 수 있다. 이 연구에서는 각 화자에 대한 휴지 시간 분석을 통하여 휴지 시간의 변화를 살펴보았다.

아래 Table 4.는 화자별 총 휴지 시간, 총 휴지 비율, 총 휴지 횟수, 평균 휴지 시간, 휴지 간 평균 조음 시간 및 평균 조음 음절수이다.

Table 4. duration and syllables of articulation across-speakers

Speaker	Total duration of pause	Ratio of pause	Number of pauses	Average duration of pause	Average duration of articulation	Average syllables of articulation
Ann	8.83sec	13.2%	21	0.42sec	2.63sec	16.3
Gyuri	8.53sec	11.7%	24	0.36sec	2.58sec	14.3
Yumi	8.53sec	11.7%	24	0.36sec	2.58sec	14.3
nVoice	6.76sec	10.0%	30	0.23sec	1.97sec	11.5

각 항목들에 대한 측정 방법은 다음과 같다. 다음의 측정 방법은 신지영(2018, pp. 99-100)에서 사용한 측정 방법을 사용하였다.

● 측정 방법

- 총 휴지 시간: 총 발화 시간 내에서 묵음이 관찰되는 구간의 총 시간
- 총 휴지 비율: 총 발화 시간 중 휴지 시간이 차지하는 비율
- 평균 휴지 시간: 총 휴지 시간을 총 휴지 횟수로 나눈 시간
- 휴지 간 평균 조음 시간: 휴지가 끝나는 지점부터 다음 휴지가 시작되는 지점 사이의 평균 조음 시간
- 휴지 간 평균 조음 음절 수: 휴지가 끝나는 지점부터 다음 휴지가 시작되는 지점 사이의 평균 조음 음절 수

휴지 시간을 측정한 결과 아나운서의 총 휴지 시간이 8.83초로 가장 길었고 총 휴지 비율도 13.2%로 가장 높았다. 규리와 유미는 총 휴지 시간과 총 휴지 비율이 8.53초와 11.7%로 같았으며 nVoice는 총 휴지 시간이 6.76초로 가장 짧았고 총 휴지 비율도 10.0%로 가장 낮았다. 그러나 총 휴지 횟수는 nVoice가 30회로 가장 많았고 규리와 유미는 24회이었으며 아나운서는 21회였다. 아나운서는 휴지 횟수는 가장 적었지만 한 번 휴지할 때 평균 0.42초를 쉬었고 규리와 유미는 0.36초를 쉬었으며 nVoice는 0.23초를 쉬었다. 이처럼 아나운서의 평균 휴지 시간이 가장 긴 이유는 아나운서는 휴지를 할 때 숨을 들이쉬면서 다음 발화를 위한 호흡을 하기 때문이

다. 아나운서는 휴지 간 조음 시간도 가장 길었고 휴지 간 조음 음절 수 역시 가장 많았다. 즉 아나운서는 한 번 휴지를 할 때 숨을 들이쉬면서 상대적으로 긴 시간 동안 휴지를 하고 휴지 이후에는 숨을 내쉬면서 상대적으로 긴 시간 동안 조음하는 것을 확인하였다. 반면에 음성합성기는 숨을 들이쉬거나 내쉬지 않기 때문에 아나운서와 비교해 짧게 휴지 시간을 설정하고 휴지 간 조음 시간 역시 짧게 설정하였다.

이어서 각 화자별로 휴지 시간에 변화가 있는지를 살펴보았다. 아래 Figure 9.는 아나운서의 휴지를 나타낸 것이다. 아나운서의 휴지 위치와 휴지 시간은 Figure 9. 아래와 같다. 휴지는 #으로 표시하였고 들숨을 동반한 휴지는 #B로 표시하였다.

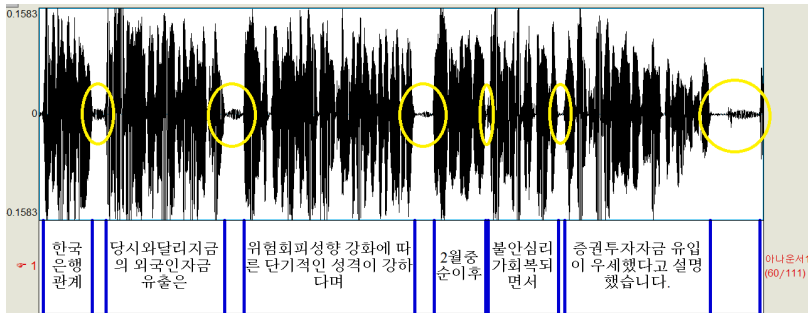


Figure 9. the pause of the announcer

한국은행관계자는 #B(0.29sec) 당시와달리지금의외국인자금유출은
#B(0.42sec) 위험회피성향강화에따른단기적인성격이강하다며
#B(0.42sec) 2월중순이후 # (0.06sec) 불안심리가회복되면서 # (0.14sec)
증권투자자금유입이우세했다고설명했습니다 #B(1.05sec)

아나운서는 휴지의 시간에 차이를 두어 끊어 읽기의 단위인 구나 절의 길이가 길 경우에는 휴지 시간이 길었고 끊어 읽기의 단위가 짧을 경우에는 휴지 시간이 짧았다. 또한 숨을 들이쉬는 휴지

에서는 비교적 휴지가 긴 반면 숨을 들이쉬지 않는 휴지에서는 휴지가 짧았다. 즉 아나운서의 휴지 시간은 끊어 읽기 단위의 길이와 숨을 들이쉬는 들숨 여부에 따라 차이가 있었다.

아래 Figure 10.은 규리의 휴지를 나타낸 것이다. 규리의 휴지 위치와 휴지 시간은 Figure 10. 아래와 같다.

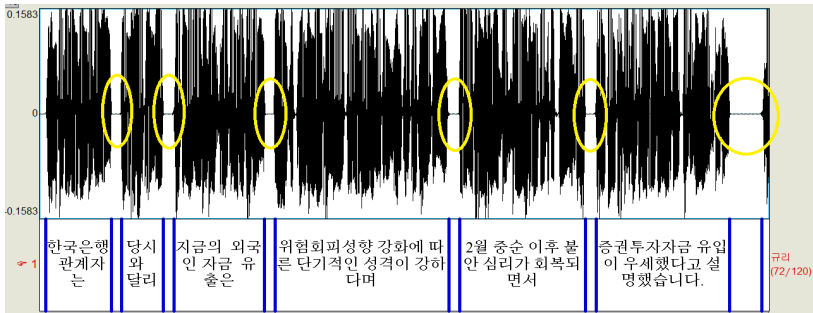


Figure 10. the pause of Gyu-ri

한국은행관계자는 #(0.25sec) 당시와달리 #(0.25sec) 지금의외국인자금
유출은 #(0.25sec) 위험회피성향강화에따른단기적인성격이강하다며
#(0.25sec) 2월중순이후불안심리가회복되면서 #(0.25sec) 증권투자자금
유입이우세했다고설명했습니다. #(0.75sec)

규리는 아나운서와 달리 들숨을 동반한 휴지가 발견되지 않았고 휴지 시간에도 차이가 없어 끊어 읽기 단위의 길이와 상관없이 문장 내 휴지 시간은 0.25초, 문장과 문장 간의 휴지 시간은 0.75초로 휴지 시간이 일정하였다.

아래 Figure 11.은 nVoice의 휴지를 나타낸 것이다. nVoice의 휴지 위치와 휴지 시간은 Figure 11. 아래와 같다.

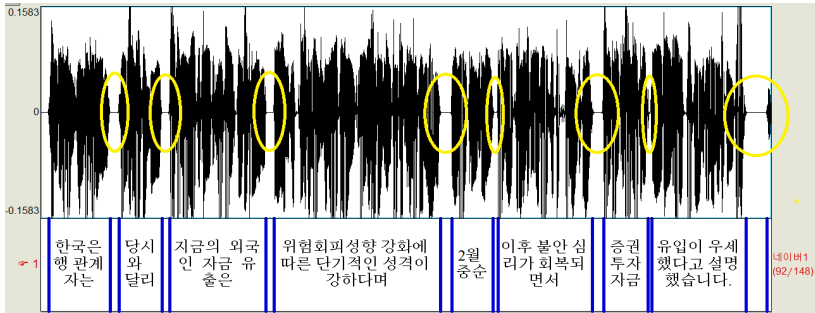


Figure 11. the pause of nVoice

한국은행관계자는 # (0.19sec) 당시와달리 # (0.17sec) 지금의외국인자금
유출은 # (0.17sec) 위험회피성향강화에따른단기적인성격이강하다며
(0.24sec) 2월중순 # (0.10sec) 이후불안심리가회복되면서 # (0.24sec) 증
권투자자금 # (0.07sec) 유입이우세했다고설명했습니다 # (0.45sec)

nVoice의 휴지 위치는 규리와 다른데, 이는 음성 합성 엔진인 nVoice와 VoiceText가 문장 내에 휴지를 주는 경계를 다르게 분석하였기 때문이다. 또한 nVoice는 규리와 달리 문장 내 휴지 시간이 동일하지 않고 아나운서와 같이 문장 내 짧은 휴지와 긴 휴지의 시간 차이를 두었다. 즉 문장 내의 끊어 읽기 단위의 길이가 짧은 경우에는 휴지 시간이 짧았고 끊어 읽기 단위의 길이가 길 경우에는 휴지 시간이 길었는데, 이는 nVoice가 문장 내에 경계를 부여할 때 끊어 읽기에 해당하는 강세구³⁾ 경계와 휴지에 해당하는 억양구⁴⁾ 경계로 나누어 경계를 부여하기 때문이다.

그런데 nVoice의 끊어 읽기에서 오류가 발견되었다. 아래 Figure

- 3) 하나 이상의 단어들이 모여서 이보다 상위의 운율 단위인 강세구(AP, Accentual Phrase)를 형성하게 된다. 강세구는 발화를 할 때 짧게 숨을 끊어 쉬는 단위이다(신지영, 차재은, 2003, pp. 139-147).
- 4) 하나 이상의 강세구들이 모여서 이보다 상위의 운율 단위인 억양구(IP, Intonational Phrase)를 형성하게 된다. 억양구의 마지막 음절 뒤에 휴지가 따라오게 된다(신지영, 차재은, 2003, pp. 148-153).

12.에서 ‘2월 중순 이후’ 다음에 끊어 읽기를 해야 하는데, ‘2월 중순’과 ‘이후’ 사이를 잘못 끊어 읽은 후 ‘이후불안심리가회복되면서’를 이어서 낭독하였고 ‘증권 투자 자금 유입이’ 다음에 끊어 읽기를 해야 하는데, ‘증권 투자 자금’과 ‘유입이’ 사이를 잘못 끊어 읽은 후 ‘유입이우세했다고설명했습니다’를 이어서 낭독하였다.

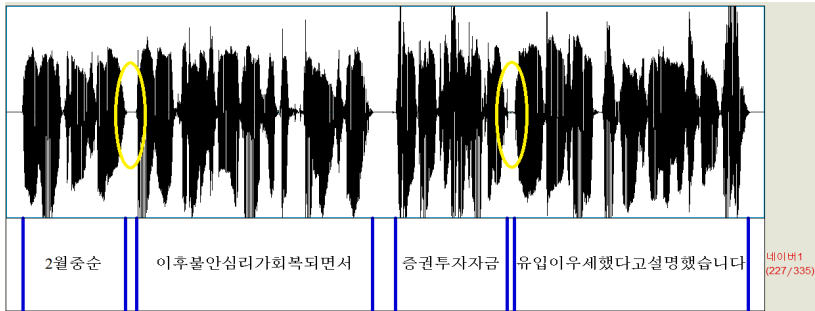


Figure 12. nVoice's error in reading by punctuating

VoiceText 역시 끊어 읽기에서 오류가 발견되었다. 아래 Figure 13.에서 규리는 ‘외국인 주식’과 ‘투자 자금이’ 사이를 잘못 끊어 읽은 후 ‘투자자금이사년여만에’를 이어서 낭독하였다.

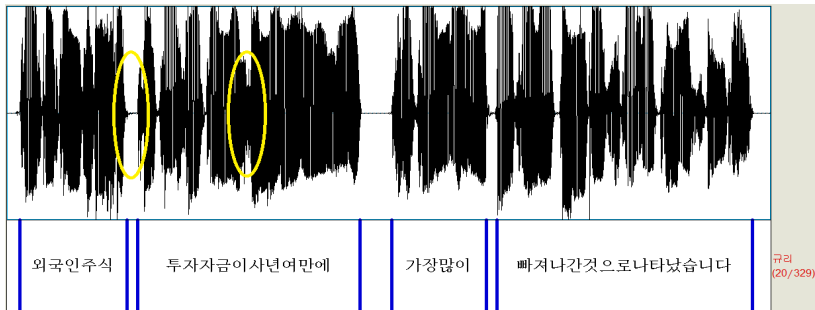


Figure 13. Gyuri's error in reading by punctuating

음성합성기에서는 텍스트가 입력되면 형태소 분석기에서 어휘 사전을 이용하여 형태소 단위로 분리하고 품사를 추정한 다음 구문 분석기에서 문장의 구문 구조를 분석한다. 이후 문장 내의 각 어절의 품사 정보를 이용하여 어절의 경계를 자동으로 추정한다(고락환, 2005, pp. 26-31). 그런데 문장의 형태소 분석과 구문 구조 분석이 잘못될 경우 어절 경계를 잘못 설정하는 오류가 발생하게 된다. nVoice는 ‘2월 중순 이후’의 부사구와 ‘증권 투자 자금 유입’의 명사구를 각각 하나의 단위로 인식하지 못하고 어절 경계를 중간에 잘못 부여하였고 VoiceText는 ‘외국인 주식 투자 자금’을 하나의 단위로 인식하지 못해 잘못 끊어 읽고 주어를 끊어 읽지 않고 이어서 읽는 등 어절 경계 설정의 오류가 있었다.

4. 전달 속도

이 연구에서는 각 화자별로 어절의 조음 속도를 측정해 전달 속도에서 어떠한 차이가 있는지를 살펴보았다. 각 음성파형의 조음 음절 수와 조음 시간을 측정한 다음, 조음 음절 수를 조음 시간으로 나누어 조음 속도를 계산하였다.

먼저 화자별 발화 시간과 발화 속도⁵⁾, 조음 시간과 조음 속도⁶⁾를 측정하였다. 아래 Table 5.에서 휴지 시간을 포함한 발화 속도를 보면 아나운서는 1분당 322음절을 낭독하였고 규리와 유미는 294음절을 낭독하였으며 nVoice는 316음절을 낭독하였다. 아나운서와 nVoice의 전달 속도가 VoiceText보다 더 빨라 속도감에 있어서 더

5) 발화 시간은 화자가 발화를 위해 조음을 시작한 시점부터 발화를 종료한 시점 사이의 시간으로서 발화 중간의 휴지를 포함한 시간이다. 이 연구에서 발화 속도는 휴지 시간을 포함한 시간당 발화 음절 수로 계산하였다.

6) 조음 시간은 발화 시간 중 조음이 이루어진 구간의 시간으로서 발화 중간의 휴지를 제외한 시간이다. 이 연구에서 조음 속도는 휴지 시간을 제외한 시간당 조음 음절 수로 계산하였다.

유창하다는 느낌을 주었다. 휴지 시간을 제외한 조음 속도는 아나운서가 1초당 6.2음절로 가장 빨랐고 nVoice 5.9음절, 규리와 유미 5.6음절순이었다.

Table 5. duration and pace of articulation across-speakers(“syl” means syllables)

Speaker	Duration of total speech	Speech pace	Duration of total articulation	Articulation pace
Announcer	66.68sec	322syl/min	57.85sec	6.2syl/sec
Gyuri	72.99sec	294syl/min	64.46sec	5.6syl/sec
Yumi	72.99sec	294syl/min	64.46sec	5.6syl/sec
nVoice	67.92sec	316syl/min	61.16sec	5.9syl/sec

한편 화자별로 조음 속도에 변화가 있는지를 살펴보았다. 아래 Figure 14.는 ‘달러당 1,082원 80전을 기록했습니다’를 낭독한 아나운서의 음성파형과 조음 시간을 나타낸 것이다. 아나운서는 문장 내에서 중요한 정보를 가지고 있는 어절인 ‘달러당 천팔십이원 팔십전을’을 낭독할 때는 아나운서의 평균 조음 속도인 초당 6.2음절보다 다소 느린 초당 5.6음절(12음절/2.124초)의 속도로 낭독한 반면 정보를 가지고 있지 않은 어절인 ‘기록했습니다’를 낭독할 때는 다소 빠른 초당 7.8음절(6음절/0.766초)의 속도로 낭독하였다.

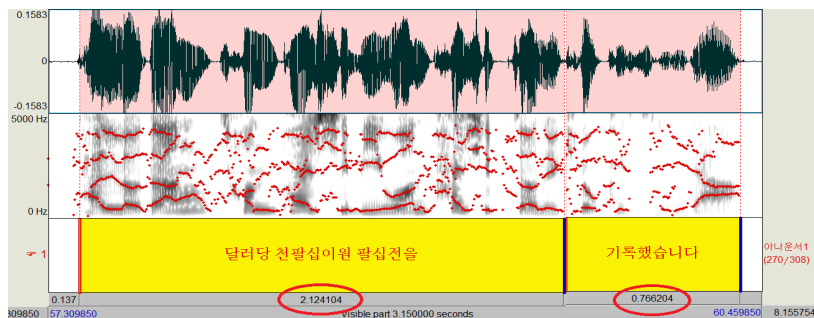


Figure 14. change of articulation pace by the announcer

그러나 규리와 유미, nVoice는 중요한 정보를 가지고 있는 어절과 그렇지 않은 어절을 낭독할 때의 조음 속도 차이가 크지 않았다. Figure 15.에서 규리가 ‘달려당 천팔십이원 팔십전을’을 낭독할 때의 조음 속도는 초당 5.4음절(12음절/2.226초)이었고 ‘기록했습니다’를 낭독할 때의 조음 속도는 초당 6.6음절(6음절/0.906초)로, 아나운서와 비교해 조음 속도의 차이가 작았다.

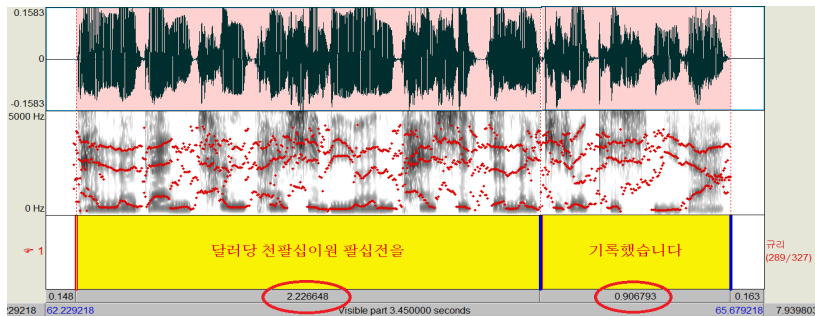


Figure 15. change of articulation pace by Gyuri

아래 Table 6.은 실험 뉴스의 문장들을 끊어 읽기 단위로 나누어 각 화자들의 조음 속도 변화를 나타낸 것이다. 아나운서는 각 문장 내에서 정보를 가지고 있는 어절들을 낭독할 때는 속도를 느리게 하여 정보의 내용을 강조한 반면 문장의 마지막 서술어인 ‘나타났습니다’, ‘기록했습니다’ 등 정보를 가지고 있지 않은 어절들을 낭독할 때는 속도를 다소 빠르게 하였다. 반면에 규리와 유미, nVoice는 평균 조음 속도인 초당 5.6음절과 5.9음절에서 큰 차이 없이 모든 어절들을 거의 일정한 속도로 낭독하였다.

Table 6. change of articulation pace across-speakers

Speaker	미국의금리인상과관련해	세계금융시장이흔들린지난달	외국인주식투자자금이
Announcer	6.7syl(1.63sec)	6.5syl(1.99sec)	6.6syl(1.51sec)
Gyuri	5.2syl(2.10sec)	5.4syl(2.42sec)	6.0syl(1.67sec)
Yumi	5.5syl(2.00sec)	5.4syl(2.40sec)	6.2syl(1.60sec)
nVoice	6.2syl(1.76sec)	6.1syl(2.12sec)	6.2syl(1.62sec)
Speaker	사년여만에가장많이빠져나간것으로	나타났습니다	한국은행이지난팔일발표한
Announcer	6.4syl(2.49sec)	8.1syl(0.74sec)	6.6syl(1.80sec)
Gyuri	5.8syl(2.75sec)	6.9syl(0.86sec)	5.4syl(2.23sec)
Yumi	5.8syl(2.74sec)	6.8syl(0.88sec)	5.4syl(2.22sec)
nVoice	6.4syl(2.51sec)	6.5syl(0.92sec)	6.2syl(1.93sec)
Speaker	이천십팔년이월중	국제금융외환시장동향을보면	지난달국내에서빠져나간
Announcer	5.1syl(1.57sec)	6.0syl(2.17sec)	6.8syl(1.60sec)
Gyuri	5.0syl(1.61sec)	5.5syl(2.37sec)	5.5syl(1.98sec)
Yumi	5.0syl(1.59sec)	5.1syl(2.54sec)	5.9syl(1.84sec)
nVoice	5.2syl(1.53sec)	5.5syl(2.36sec)	6.4syl(1.73sec)
Speaker	외국인주식자금은	삼십육억삼천만달러로	이천십삼년유월이후최대규모를
Announcer	6.2syl(1.28sec)	5.4syl(1.84sec)	5.8syl(2.43sec)
Gyuri	5.9syl(1.34sec)	5.0syl(2.00sec)	5.3syl(2.64sec)
Yumi	5.6syl(1.41sec)	5.5syl(1.82sec)	5.4syl(2.57sec)
nVoice	5.8syl(1.38sec)	6.0syl(1.67sec)	5.7syl(2.47sec)
Speaker	기록했습니다		
Announcer	8.0syl(0.75sec)		
Gyuri	6.6syl(0.90sec)		
Yumi	6.6syl(0.90sec)		
nVoice	6.9syl(0.86sec)		

이처럼 음성합성기의 합성음이 조음 속도의 변화가 적은 이유는 음성 DB에 저장할 음편들을 녹음할 때 성우들이 발화 속도가 일정하게 유지되도록 일관성을 유지하여 녹음하기 때문이다.

5. 음장

장음을 단음보다 길게 발음하는 음장의 변화는 음성파형과 스펙트로그램(spectrogram) 분석을 통해 확인할 수 있다. 이 연구에서는 각 화자별로 단어의 첫음절 조음 시간을 측정해 장음과 단음의 음장 시간에 어떠한 차이가 있는지를 살펴보았다.

Figure 16.은 아나운서가 ‘4년여 만에’를 발음한 음성파형과 스펙트로그램이다. 첫음절인 ‘사’는 장음으로 길게 발음해야 하는데, 아나운서는 257msec의 시간으로 발음해 ‘년(131msec)’, ‘여(142msec)’, ‘만(149msec)’, ‘에(101msec)’보다 길게 발음하였다.

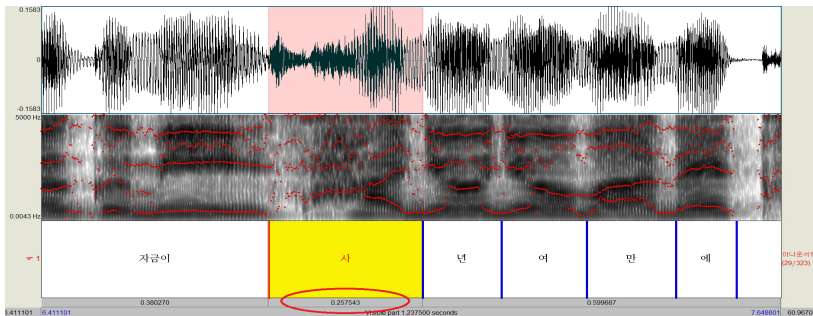


Figure 16. vowel length of articulation by the announcer

한편 규리와 유리, nVoice는 첫음절인 ‘사’와 이후의 음절들을 큰 시간 차이 없이 발음하였다. Figure 17.에서 규리는 ‘사’를 181msec의 시간으로 발음해 ‘년(148msec)’, ‘여(187msec)’, ‘만(182msec)’, ‘에(194msec)’와 큰 시간 차이가 없었다.

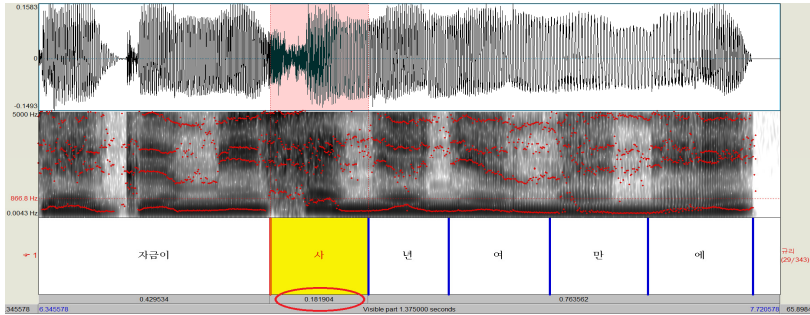


Figure 17. vowel length of articulation by Gyuri

다음 Table 7.은 실험 뉴스 문장 전체의 29개 장음 단어와 67개 단음 단어의 첫음절을 화자별로 발음한 평균 시간과 장단음의 음장 상댓값을 나타낸 것이다. 아나운서는 장음 단어의 첫음절을 평균 190msec로 가장 길게 발음하였고 규리 180msec, 유미 179msec, nVoice 175msec 순으로 발음하였다. 또한 아나운서는 단음 단어의 첫음절을 평균 144msec로 발음하였고 규리 178msec, 유미 177msec, nVoice 170msec의 시간으로 발음하였다. 화자별 장단음의 음장 상댓값을 보면 아나운서는 단음에 비해 장음을 1.319배 길게 발음하였으나 규리와 유미, nVoice는 장단음 간의 음장 차이 없이 비슷한 시간으로 발음하였다.

Table 7. length of long and short vowels across-speakers

Speaker	Length of long vowels	Length of short vowels	Relative value of length
Announcer	190msec	144msec	1.319
Gyuri	180msec	178msec	1.011
Yumi	179msec	177msec	1.011
nVoice	175msec	170msec	1.029

이처럼 음성합성기의 장단음 간 음장 시간에 차이가 없는 것은 성우들이 음성 DB에 저장할 음편들을 녹음할 때 장단음 간 음장 시간에 차이를 두지 않고 일정한 길이로 녹음하였기 때문이다.

6. 발음 개구도

모음의 발음에서 입을 벌리는 정도인 개구도는 포먼트(formant) 값을 통해 확인할 수 있다. 포먼트는 성도(聲道)의 공명(共鳴) 특성을 음향적인 수치로 구한 값을 말한다. 포먼트는 소리를 낼 때 턱과 혀 등의 조음기관의 위치를 파악해 특징을 잡아내기 위해 찾는 것으로 컴퓨터 알고리즘에 의해 자동으로 계산해 붉은 색 띠의 형태로 나타낸다. 아래로부터 차례로 제1포먼트, 제2포먼트 등으로 부르는데, 제1포먼트는 턱의 열림 정도를 나타내어 입을 많이 벌릴수록 높아지고 입을 다물수록 낮아진다(양병곤, 2010, p. 19).

이 연구에서는 각 화자별로 저모음 발음의 제1포먼트 값을 측정해 개구도에서 어떠한 차이가 있는지를 살펴보았다. 아래 Figure 18은 ‘4년여 만에 가장 많이 빠져나간 것으로 나타났습니다’를 발음한 아나운서의 음성파형과 포먼트 곡선을 나타낸 것이다. 맨 아래에 있는 곡선이 제1포먼트 곡선이다. 제1포먼트는 턱의 오르내림을 반영하여 입을 많이 벌릴수록 높아지고 입을 다물수록 낮아지는데, ‘사, 년, 여, 만, 가, 장, 많, 빠, 저, 나, 간, 것, 나, 타, 났, 다’와 같이 ㅏ, ㅑ, ㅓ 등의 저모음이 포함된 음절을 발음할 때 제1포먼트 곡선이 높고, ‘이, 으, 습, 니’와 같이 ㅡ, ㅣ 등의 고모음이 포함된 음절을 발음할 때 제1포먼트 곡선이 낮은 것을 확인할 수 있다. 그림에 나타난 구간에서 아나운서의 저모음 음절 발음의 제1포먼트 평균값은 820Hz로 측정되었다.

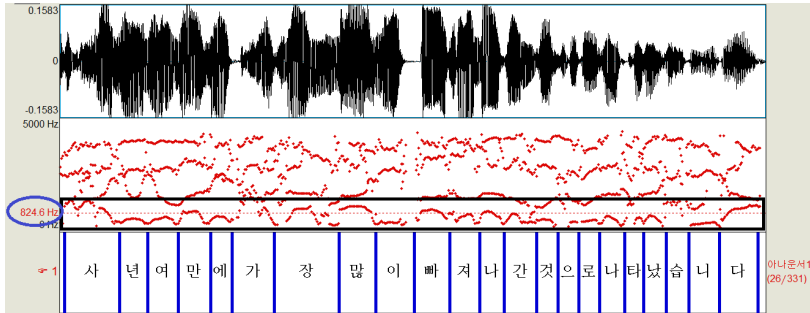


Figure 18. the first formant curve of the announcer

이에 비해 규리와 유리, nVoice의 저모음 음절 발음의 제1포먼트 평균값은 654Hz, 650Hz, 672Hz로 측정되어 아나운서보다 낮게 나타났다. 아래 Figure 19.에서 규리의 제1포먼트 곡선이 아나운서보다 낮게 형성되어 있음을 확인할 수 있는데, 이는 규리가 아나운서보다 입을 덜 벌리면서 저모음들을 발음하고 있음을 나타내고 있다.

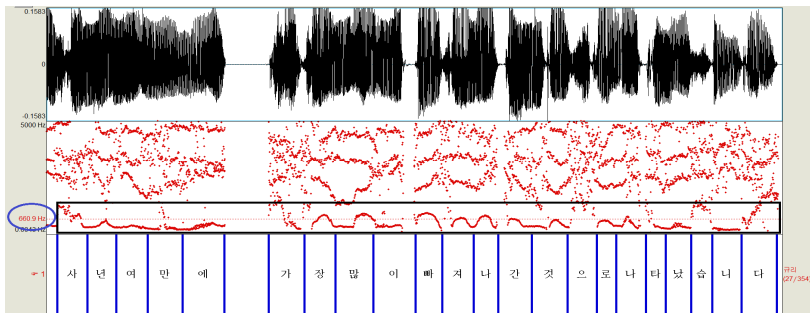


Figure 19. the first formant curve of Gyuri

아래 Table 8.은 화자별로 실험 뉴스 문장 전체에서 저모음이 포함된 음절들을 발음할 때의 제1포먼트 평균값을 나타낸 것이다. 분석 결과 입을 가장 크게 벌려 발음해야 하는 ㅏ 모음이 포함된 77음절의 제1포먼트 평균값은 아나운서가 가장 높았고 ㅏ, ㅑ, ㅓ,

조의 억양을 구사하는 등 억양에 변화를 주었다. 음성합성기는 음편을 녹음한 성우의 낭독에 따라 음높이 변동 폭에 차이가 있었고 아나운서와 달리 전체적으로 상승조 혹은 하강조 억양을 구사하는 특징을 보였다.

둘째, 아나운서는 가장 폭넓게 음의 강도를 변화시키며 뉴스를 낭독하였고 정보를 가지고 있는 단어들은 일정한 강도를 유지하며 낭독한 반면 정보를 가지고 있지 않은 단어를 낭독할 때는 강도를 줄여 낭독하였다. 그러나 음성합성기는 문장 내의 모든 단어들을 비교적 일정한 강도로 낭독하였다.

셋째, 아나운서는 숨을 들이쉬는 휴지와 끊어 읽기 단위의 길이가 긴 휴지에서는 비교적 긴 휴지를 둔 반면 숨을 들이쉬지 않는 휴지와 끊어 읽기 단위의 길이가 짧은 휴지에서는 짧은 휴지를 두었고 nVoice는 문장 내의 끊어 읽기 단위의 길이에 따라 휴지 시간에 변화를 주었으나 VoiceText는 문장 내 끊어 읽기 단위의 길이에 상관없이 일정한 시간 간격으로 휴지를 두었다. 또한 음성합성기에서는 하나의 단위로 이어서 읽어야 할 구를 중간에 잘못 끊어 읽거나 끊어 읽어야 할 두 개의 구를 이어서 읽는 등 끊어 읽기에서 오류가 발견되었다.

넷째, 아나운서는 문장 내에서 중요한 정보를 가지고 있는 단어를 낭독할 때는 속도를 느리게 하여 길이를 증가시키고 정보를 가지고 있지 않은 단어들을 낭독할 때는 속도를 빠르게 하여 길이를 감소시키는 등 변화를 주었으나 음성합성기는 각 어절별로 조금 속도의 차이가 크지 않았고 거의 일정한 속도로 낭독하였다.

다섯째, 아나운서는 장단 모음 간 음장 시간에 차이를 두어 변별력 있게 낭독한 반면 음성합성기는 장단 모음 간 음장 시간에 차이를 두지 않고 낭독하였다.

여섯째, 아나운서는 입을 크게 벌려야 하는 ㅏ, ㅑ, ㅕ 등의 저모음을 발음할 때 음성합성기보다 대부분 입을 크게 벌려 발음하였으나 음성합성기는 아나운서보다 상대적으로 입을 작게 벌려 발음

하였다.

정리하면 아나운서는 유사언어에 변화를 주면서 유사언어 요소들을 조절하여 뉴스를 낭독하는 데 비해 음성합성기는 대부분의 유사언어 요소에 변화를 주지 않고 일정하게 뉴스 문장을 낭독하는 경향을 보였다.

음성 합성 기술은 음편들을 사전에 사람이 녹음하여 음성 DB에 저장하여 두고 텍스트가 입력되면 텍스트의 내용과 일치하는 음편들을 불러와 연결하는 과정에서 억양과 강도, 휴지의 길이 등의 운율을 별도로 처리하지 않고 이미 운율이 실려 있는 음편을 선택하여 합성하는 방식이기 때문에 화자가 사전에 음편들을 어떻게 녹음하느냐에 따라 합성음의 품질이 결정된다. 따라서 음성 DB의 음편을 녹음할 화자를 선정할 때 억양, 강도, 끊어 읽기, 전달 속도, 음장, 발음의 개구도 등 유사언어의 조절 능력이 뛰어난 화자를 선정하는 것이 중요하다고 하겠다. 또한 음성합성기의 끊어 읽기에서 오류가 발생하지 않도록 텍스트가 음성합성기에 입력되었을 때 문장의 형태소와 구문 구조를 분석하는 분석기의 성능을 보다 향상시켜 어절 경계를 정확히 설정하도록 해야 한다.

이 연구는 아나운서와 음성합성기의 뉴스 낭독 음성을 음성 분석 프로그램을 이용하여 분석함으로써 유사언어의 차이를 실증적으로 규명하였다는 데 의의가 있다. 앞으로 아나운서와 음성합성기의 유사언어 차이가 실제 수용자의 뉴스 이해도와 호감도, 공신력 평가, 재청취 의도 등에 어떠한 영향을 미치는지에 대한 수용자 인식 조사가 후속 연구로 이어져야 할 것이다. 음성합성기의 합성음은 사람의 목소리를 짜깁기한 것인데, 과연 이러한 합성음을 수용자가 청취하였을 때 뉴스의 내용을 잘 이해할 수 있을지, 호감도나 공신력을 느낄 수 있을지, 또한 이후에도 합성음으로 뉴스를 재청취할 생각이 있을지에 대한 수용자 인식 조사를 통해 합성음의 자연성과 명료도를 평가할 수 있을 것이다.

참고문헌

- 고락환. (2005). 코퍼스 기반 한국어 TTS 시스템의 구현 및 성능평가에 관한 연구. 석사학위논문, 원광대학교, 익산. Ko, L. H. (2005). A study of implementation and evaluation of corpus based Korean TTS system. Master's Thesis, Wonkwang University, Iksan.
- 남도현, 임성수, 윤보람, 최홍식, 유애리. (2010). 아나운서와 일반인 화자의 음성 차이 분석에 대한 실험 연구. **KBS 한국어연구논문**, 60, 66-103.
- Nam, D. H., Rheem, S. S., Yun, B. R., Choi, H. S., & Yoo, A. R. (2010). Voice comparisons between announcers and ordinary people through reading aloud the reading-style sentences. *Journal of KBS Korean Study*, 60, 66-103.
- 박경희. (2003). 방송언어의 운율적 자질에 대한 고찰: 고저장단의 상대값 비교를 중심으로. **KBS 한국어연구논문**, 54, 3-56. A study on the rhythmic characteristics of broadcasting language based on differences of the vowel length between short vowels and long vowels. *Journal of KBS Korean Study*, 54, 3-56.
- 박경희. (2004). 유사 언어가 방송 메시지 전달에 미치는 영향에 관한 연구: 뉴스 전달 속도 분석을 중심으로. 석사학위논문, 성균관대학교, 서울.
- Park, K. H. (2003). A study on the influence of paralanguage on the transmission of broadcast message: Based on an analysis of news-reading speed. Master's Thesis, Sungkyunkwan University, Seoul.
- 신지영. (2018). 언어 수행에서의 호흡과 기억: 호흡 단위와 휴지 단위의 양적 분석 결과를 바탕으로. **한국어학**, 79(5), 91-116. Shin, J. Y. (2018). Breath and memory in speech based on quantitative analysis of breath groups and pause units in Korean. *Korean Linguistics*, 79(5), 91-116.
- 신지영, 차재은. (2003). **우리말 소리의 체계: 국어 음운론 연구의 기초를 위하여**. 서울: 한국문화사. Shin, J. Y. & Cha, J. E. (2003). *The system of Korean sound: For the basis of the study of Korean phonology*. Seoul: Hankukmunhwa.
- 양병곤. (2002. 11.). 목소리는 제2의 주민등록증. **과학동아**, 2002(11), 90-95.
- Yang, B. G. (2002. 11.). Voice is the second resident registration card. *Donga Science*, 2002(11), 90-95.

- 양병곤. (2010). **Praat(프라트)를 이용한 음성분석의 이론과 실제**. 부산: 만수출판사. Yang, B. G. (2010). *Theory and practice of voice analysis using Praat*. Busan: Mansu Publisher.
- 이도영. (2002). 음성 언어 교육과 문화 창조. **한국초등국어교육학보**, 20, 25-45. Lee, D. Y. (2002). Voice language education and culture creation. *Journal of Korean elementary korean language education*, 20, 25-45.
- 이종석, 박기태, 이준우. (2014). 시각 및 언어장애인을 위한 음성합성 기술의 현황. **전자공학회지**, 41(3), 28-39. Lee, J. S., Park, G. T., & Lee, J. W. (2014). Current status of speech composition technologies for the blind and language disabled. *The Magazine of the IEEK*, 41(3), 28-39.
- 이호영. (1991). 한국어의 리듬: 숨쉬기와 끊어 읽기를 중심으로. **KBS 한국어연구논문**, 28, 3-43. Lee, H. Y. (1991). The rhythm of Korean: Focused on breathing and reading by punctuating. *Journal of KBS Korean Study*, 28, 3-43.
- 전은주. (1991). 한국어 초점(Focus)의 실험음성학적 연구. 석사학위논문, 서울대학교, 서울. Jeon, E. J. (1991). A study on the experimental phonetics of the focus of Korean language. Master's Thesis, Seoul National University, Seoul.
- 정동빈. (1997). 뉴스 낭독 기법. **KBS 한국어연구논문**, 47, 29-71. Jeong, D. B. (1997). A news reading technique. *Journal of KBS Korean Study*, 47, 29-71.
- 조동욱. (2009). 음성 특징 추출을 통한 방송 진행자의 전달성 비교 분석. **KBS 한국어연구논문**, 59, 36-73. Cho, D. U. (2009). The communicability observations of broadcasting programs MC by extracting voice features. *Journal of KBS Korean Study*, 59, 36-73.
- 조철우, 이상호, 김수진. (2005). 한국어 합성음 평가 가이드라인 시안. **한국정보과학회 영남지부 학술발표논문집**, 13(1), 101-105. Jo, C. W., Lee, S. H., & Kim, S. J. (2005). Evaluation guidelines of Korean synthetic speech. *Communications of the Korean Institute of Information Scientists and Engineers*, 13(1), 101-105.
- 지민제, 한성숙, 김윤기. (1996). 한국어 말소리의 길이. **KBS 한국어연구논문**, 45, 65-102. Zhi, M. J., Han, S. S., & Kim, Y. G. (1996). The length of Korean speech. *Journal of KBS Korean Study*, 45, 65-102.

원고접수일: 2018년 6월 28일

원고심사일: 2018년 8월 11일

게재확정일: 2018년 8월 14일