

자연스러운 여성 합성음을 위한 지속시간 규칙에 관한 연구

최영익, 권철홍
대전대학교 정보통신공학과

The Rule of Duration Variation For Natural Female Synthetic Speech

Young-Ig Choi, Chul-Hong Kwon
Dept. of Information & Communication Eng., Taejon University
e-mail : chkwon@dragon.taejon.ac.kr

요 약

합성음의 자연성은 운율(prosody)과 관계가 있으며, 운율은 지속시간(duration), 세기(intensity), 피치(pitch)의 3가지 요소가 어우러져 이루어진다. 본 연구에서는 한국어 여성 음에 나타나는 지속시간 패턴을 분석하여 지속시간의 규칙을 정리함을 목적으로 한다. 본 연구에서는 각 음소(자음, 모음)의 고유 지속시간과 단어내의 음절 위치, 인접음소의 영향, 구와 절의 경계의 영향에 따른 지속시간의 변화를 조사하여 지속시간 규칙을 정립하였다. 청취 실험 결과, 본 지속시간 규칙이 합성음의 자연성을 향상시켰다는 것을 보여준다.

1. 서 론

요즈음 음성합성에 대한 관심이 커지면서 학교나 연구소, 기업 등에서 음성합성에 대한 연구가 활발히 진행되고 있다. 음성합성의 초기에는 합성음이 인간에게 정확하고 바르게 전달되는가, 즉 인간이 합성음을 듣고 바르게 이해할 수 있는가 하는 합성음의 명료도(clarity)에 중점을 두었다. 그러나 이것은 의미 전달은 가능하더라도 단순한 기계음에 불과하다. 합성음의 궁극적인 목표가 인간 언어라면 명료도 외에 자연성(naturalness)이 고려되어야 한다. 자연성은 운율(prosody)과 관계가 있으며, 운율은 지속시간(duration), 세기(intensity), 피치(pitch)의 3가지 요소가 어우러져 이루어진다[1,2,3]. 우리말의 경우, 지속시간이 큰 비중을 차지하기 때문에 지속

시간에 대해 체계적이고 광범위한 연구가 필요한 실정이다. 지속시간 패턴은 상황과 문맥에 따라 다양하게 변한다. 합성음에 대화체의 생생한 현장감 또는 낭독체의 차분함을 줄 수 있는 것은 지속시간이다.

운율을 조절하는 요소로서 지속시간은 매우 중요한 역할을 한다. 음소의 지속시간은 운율에 영향을 주어 합성음의 자연스러움을 결정할 뿐만 아니라 발화 속도를 결정할 수 있으므로 중요한 요소다[4,5]. 또한 자음의 경우 그 지속시간이 음가에 미치는 영향은 지대하므로 그 중요성은 더욱 커진다고 할 수 있다. 음소의 지속시간은 심리적, 의미론적 요인에 따라서도 변화할 수 있으나 이러한 정보를 추출하기가 쉽지 않으므로, 본 연구에서는 한국어 여성 음에 나타나는 지속시간 패턴을 분석하여 지속시간의 규칙을 정리함을 목적으로 한다.

2. 실험 과정

2.1 실험 목표

본 연구에서는 각 음소(자음, 모음)의 고유 지속시간과 단어내의 음절 위치, 인접음소의 영향, 구와 절의 경계의 영향에 따른 지속시간의 변화를 조사하여 지속시간 규칙을 정립한다. 조사한 구체적인 사항은 다음과 같다.

- 1) 자, 모음의 고유 지속시간
- 2) 자, 모음이 단어 내에서 차지하는 위치(단어 초, 단어 중간, 단어 말)에 따른 지속시간 변화
- 3) 어절 내 음절수에 따른 음절 당 지속시간과의 관계
- 4) 인접 음소에 따른 자음, 모음의 지속시간 변화

- 5) 문장유형(평서문, 의문문, 명령문, 인용문)에 따른 지속시간 변화
- 6) 구, 절의 경계, 문장 끝에서 음절의 지속시간 변화
- 7) 구나 절의 경계, 문장 끝에서의 휴지부의 길이
- 8) 발화 속도에 따른 지속시간 변화
- 9) 경, 격음 앞 묵음의 길이

2.2 실험절차 및 방법

사실 스튜디오의 녹음실에서 457개의 문장을 여성 화자 10인이 하루에 1번씩 3일에 걸쳐 3회 발화시켰다. Kay Electrics사의 PC용 음성 분석 프로그램인 CSL(Computerized Speech Lab.) 4300B를 이용하여 10 KHz 샘플링, 16 bit 양자화 하여 A/D 변환시켰다. 윈도우 4개를 열어서 파형, 스펙트럼그램, 기본주파수(F0), 에너지가 출력되도록 한 다음 이 네 가지를 동시에 이용하여 분석하였다.

3. 실험 결과 및 토의

2.1절의 실험 목표에 따라 자, 모음의 지속시간에 대한 데이터를 추출하여 그 특징을 파악하였다. 실험결과 지속시간 변화범칙은 다음과 같다. 본 논문에서 제시한 지속시간 데이터는 여성 화자 10인이 3번에 걸쳐 발화한 값(30개 값)의 평균치이다.

3.1 자, 모음의 고유 지속시간

음소의 고유 지속시간은 같은 음성언어학적 조건에서 발음한 자료에서 측정하여야 한다[6]. 표준말 모음의 고유 지속시간을 살펴보기 위해 같은 음성 환경에서 발음한 모음의 지속시간을 측정한 결과를 보면, /에,어,이,으/ 120 msec, /아,애,오,외/ 130 msec, /우,여,예/ 140 msec, /야,얘,요,외/ 150 msec, /와,웨,웨,위,유/ 160 msec, /워/ 170 msec로 나타났다.

우리말 자음의 고유 지속시간을 측정하기 위해 /자음+모음/ 환경을 선택했다. 그리고 /모음+자음+모음/ 환경에서 유성음화 되는 자음(예, /ㄱ, ㄷ, ㅈ, ㅊ/)의 지속시간에 대해서는 3.2 음절 위치에 따른 지속시간 변화에서 언급할 것이다. /자음+모음/ 환경에서 측정한 자음의 지속시간은 다음과 같다. /ㄱ, ㄷ, ㅈ, ㅊ/ 63, /ㅋ, ㆁ, ㆁ, ㆁ/ 61, /ㄴ, ㄹ/ 92, /ㄷ, ㄷ, ㄷ/ 23, /ㄸ/ 11, /ㅃ/ 12, /ㅅ/ 140, /ㅆ/ 152, /ㅈ/ 95, /ㅊ/ 69, /ㄷ, ㄷ/ 10, /ㅎ/ 95 msec로 나타났다. 여기에서 경, 격음의 지속시간은 묵음구간을 제외한 실제 음성의 길이이고, 자음 /ㅇ/이 단어 초에 위치할 때는 음가가 없기 때문에 제외시켰다.

중성 자음에 대해서 별도로 측정을 했다. 중성 자음 중 /ㄱ, ㄷ, ㅈ/는 내파음이기 때문에 측정을 하기가 어려

워 이들을 제외하고, /ㄴ, ㄹ, ㄹ, ㄹ/에 대하여 측정한 결과는 다음과 같다. /ㄴ/ 130msec, /ㄹ/ 130msec, /ㅇ/ 150msec, /ㄷ/ 130msec이다.

3.2 음절 위치에 따른 자, 모음 지속시간 변화

음소의 길이는 어절 내 음절의 위치에 따라 영향을 받는다. 즉, 한 음절이 처음, 중간 그리고 마지막에 위치할 때 지속시간에는 차이가 있다.

자음의 경우, 첫째 음절에 위치하는 초성 자음의 지속시간은 중간, 마지막 음절에 위치할 때보다 길고, 중간과 마지막 음절인 경우는 비슷하다. 이런 현상은 /ㄱ, ㄷ, ㅈ, ㅊ, ㅋ, ㆁ, ㆁ, ㆁ/인 경우 두드러진다. /ㄱ, ㄷ, ㅈ, ㅊ/인 경우는 이들이 중간이나 마지막 음절에 나타날 때는 유성음화가 되어, 유성음화 자음은 지속시간의 감소가 동반됨을 알 수 있다. 중성 자음 /ㄴ, ㄹ, ㄹ, ㄹ/의 경우는 이와 상반되는 현상이 관찰된다. 즉, 마지막 음절의 중성 자음은 첫째나 중간 음절에 위치할 때보다 20% 더 길었다.

모음의 경우, 측정 결과가 표 1에 도시되어 있는데, 마지막 음절의 모음 지속시간이 다른 경우보다 길어지는 경향이 있음을 알 수 있다. 이는 중성 자음인 경우와 같은 현상으로, 이 두 가지 경우로부터 지속시간의 증가가 어절의 경계에서 나타나는 현상임을 파악할 수 있다.

표 1. 음절 수, 음절 위치에 따른 모음 지속시간

음절 수	음절 위치에 따른 지속시간(msec)				
	첫째	둘째	셋째	넷째	다섯째
1	296				
2	205	263			
3	130	153	233		
4	123	150	155	231	
5	120	141	143	147	228

3.3 음절수 증가에 따른 자, 모음 지속시간 변화

어절 내 음절수에 따라 자음 및 모음의 지속시간이 어떻게 변화하는가를 관찰하였다. 자음의 지속시간은 음절수가 증가함에 따라 감소하는 경향이 있다. 그러나 음절수 증가에 따라 감소하는 정도는 음절수가 3개 이상이면 현저히 줄어든다.

모음의 지속시간도 자음에서와 같이 음절수가 증가함에 따라 감소하는 경향이 있음을 표 1을 통해 알 수 있다. 또한 음절이 3음절 이상이면 감소하는 정도가 극히 미약하고, 3음절까지 감소하는 정도는 자음인 경우에 비해 훨씬 크다. 이로부터 지속시간은 자음보다는 모음에서 더 크게 변화한다는 사실을 파악할 수 있다. 이 결과

는 모음이 자음보다 탄성이 커서 모음의 지속시간 변화가 자음보다 크다는 성유나[7]의 실험 결과와 일치한다.

3.4 인접 음소에 따른 자, 모음 지속시간 변화

인접 음소가 자음에 미치는 영향을 알아보기 위해 다음 두 가지 경우를 실험했다. 앞 종성이 뒤 초성에 미치는 영향과 뒤 초성이 앞 종성에 미치는 영향이 그것이다. 대체적으로 앞 종성은 뒤 초성 자음의 지속시간을 감소시키는 효과를 발휘하는 경향이 있는데, 단지 뒤 초성이 /ㄱ, ㄷ, ㄱ, ㅈ/인 경우는 지속시간이 증가했다.

또한 뒤 초성은 앞 종성 /ㄴ, ㄹ, ㄹ, ㄹ/의 지속시간을 감소시키는 경향이 일반적으로 나타난다. 이로부터 자음이 연속하여 발음되는 경우 상호 자음이 줄어드는 경향이 있음을 알 수 있고, 특히 뒤 초성이 /ㄱ, ㄷ, ㄱ, ㅈ, ㄷ, ㅈ, ㅈ/인 경우 목음을 포함한 이들 자음의 지속시간이 길어 앞 종성의 지속시간을 현저히 감소시킨다는 사실을 파악할 수 있다.

인접 음소가 모음에 미치는 영향을 알아보기 위해, 앞 초성이 뒤 모음에 미치는 영향, 뒤 초성이 앞 모음에 미치는 영향, 뒤 종성이 앞 모음에 미치는 영향 그리고 모음 상호간의 영향 등에 관한 실험을 했다. 앞 초성의 영향은, 비음 /ㄴ, ㄹ, ㄹ/, 유음 /ㄹ/, 경음 /ㄱ, ㄷ, ㄱ/인 경우는 모음의 지속시간을 증가시키고, 격음 /ㄱ, ㄷ, ㄱ/인 경우는 감소시키는 특징이 있다. 격음인 경우 감소하는 이유는 기(aspiration)가 모음의 안정구간까지 과도적으로 전달되기 때문이다.

뒤 초성의 영향은, 경음 /ㄱ, ㄷ, ㄱ/과 격음 /ㄱ, ㄷ, ㄱ/ 앞에서 모음이 짧아지는 경향이 있다. 모음과 모음이 연속적으로 발음될 때는 (예, /아이/), 앞 모음을 짧게 뒤 모음을 길게 발음하는 경향이 있다. 뒤 종성의 영향은, 무성 받침 /ㄱ, ㄷ, ㄱ/인 경우에 비해 유성 받침 /ㄴ, ㄹ, ㄹ/이 앞 모음의 지속시간을 짧게 하는 경향을 갖는다.

3.5 문장 유형에 따른 마지막 음절의 지속시간

평서문, 예/아니오 의문문, 의문사 의문문, 명령문, 청유문 등 문장 유형에 따른 마지막 음절의 지속시간을 측정하였다. 일반적으로 문장의 마지막 어절에서 지속시간의 큰 변화를 보여주는 음절이 마지막 음절이라고 알려져 있기 때문에 이와 같은 실험을 했다. 의문문인 경우 특히 예/아니오 의문문인 경우가(250msec) 평서문(350msec), 의문사 의문문(310msec), 명령문(330msec), 청유문(320msec) 등 다른 문장 유형보다 마지막 음절의 지속시간이 짧다는 것을 알 수 있다.

3.6 구나 절의 경계, 문장 끝 마지막 음절의 지속시간 및 휴지부의 길이

사람이 발성할 때 이해도를 향상시키기 위해 또는 숨의 조절을 위해 끊어 읽기를 한다. 이러한 끊어 읽기는 청취자가 문장에 내재된 구문구조를 이해하는데 중요한 단서가 된다. 따라서 TTS 시스템에서 합성된 문장의 자연성과 이해도를 증진시키기 위해 문장내의 적절한 위치에 끊어 읽기를 해야한다. 여기에서 끊어 읽는 부분을 경계라 하고 경계에서 나타나는 현상은 경계 앞 음절의 지속시간 변화, 경계에서 휴지부(pause)의 존재, 억양의 변화 등이다. 경계 앞 음절의 지속시간은 증가하는 현상을 보였는데, 종성이 없는 경우는 모음의 지속시간이 증가했고 종성이 있는 경우는 모음보다 종성의 지속시간이 크게 증가했다.

그리고 경계에서는 휴지부가 존재했는데 휴지부는 경계의 종류에 따라 그 길이가 차이가 있었다. 경계 종류에 따른 휴지부의 길이를 보면, 구의 경계(90msec)보다는 절의 경계(450msec)에서 휴지부의 길이가 길었다. 특이한 현상은 문장 경계(250msec)에서 휴지부의 길이가 짧다는 것인데 이는 문장 경계에서 억양의 큰 내림조 현상, 새가의 큰 감소가 동반하므로 휴지부의 길이 증가 없이도 문장 끝이라는 사실을 드러낼 수 있기 때문이라 생각된다.

3.7 발화 속도에 따른 지속시간 변화

발화속도에 따른 지속시간 변화를 살펴보기 위하여 같은 문장을 느리게, 보통 속도로, 빠르게 등으로 읽게 하였다. 발화속도 변화에 따라 나타나는 현상으로는 전체 문장 길이 변화, 문장 내 휴지부의 길이 변화, 이들의 차이로 구하는 실제 음성의 길이 변화, 경계의 수 증감을 들 수 있다.

'어디선가 바보, 바보 하는 소리가 자꾸만 뒤 따라 오는 것 같았다.'는 문장에서 측정된 결과를 분석하면 다음과 같다. 문장 전체 길이는 보통 속도에서 4700msec, 느린 속도에서 6200msec, 빠른 속도에서 3700msec로 나타났다. 즉 전체 길이는 느린 속도에서 보통 속도에서보다 32%의 증가율을 보였고, 빠른 속도에서는 21%의 감소율을 보인다. 문장 내 휴지부의 길이는 보통 속도에서 250msec, 느린 속도에서 795msec, 빠른 속도에서 80msec로 나타났다. 즉 휴지부 길이는 느린 속도에서 보통 속도에서보다 218%의 증가율을 보였고, 빠른 속도에서는 68%의 감소율을 보인다. 따라서 발화속도의 변화는 실제 음성의 길이보다는 휴지부의 길이를 더 큰 비율로 변화시킨다는 사실을 알 수 있다.

발화속도에 따른 경계의 수를 살펴보면 느린 속도에서 5군데, 보통속도에서 3군데, 빠른 속도에서 1군데로 나타났다. 따라서 발화속도의 변화는 휴지부의 길이 증가와 함께 경계의 수 증가가 동반된다는 것을 알 수 있다.

3.8 경, 격음 앞 묵음의 길이

경음 /ㄱ, ㅋ, ㆁ/과 격음 /ㄱ, ㄷ, ㅌ/ 앞에서는 단립 구간이 음향적으로 묵음으로 나타난다. 이 묵음구간을 무시하고 앞 음소에 경음 또는 격음을 바로 연결하여 합성하면 이들 음가가 제대로 들리지 않기 때문에 이 묵음구간을 고려해야 한다. 그런데 /자음+모음/ 환경에서는 이 묵음구간이 발화 시작 전의 휴지부와 구별할 수가 없으므로, /모음+자음+모음/ 또는 /중성자음(ㄱ, ㄷ, ㅌ 또는 ㄴ, ㄹ, ㄴ, ㄹ)+자음+모음/ 환경에서 이 묵음구간을 측정했다. 앞 중성이 무성 자음 /ㄱ, ㄷ, ㅌ/인 경우 묵음기간이 가장 길고(74msec), 앞 중성이 유성 자음 /ㄴ, ㄹ, ㄴ, ㄹ/인 경우와(60msec) 경, 격음이 앞 모음과 바로 연결될 때(62msec) 묵음구간이 비슷하다.

4. 지속시간 변화법칙의 평가

위의 지속시간 변화규칙을 합성음에 적용한 뒤, 기존의 합성음과 비교해 자연성이 어느 정도 향상되었는가를 살펴보았다.

자연성의 평가는 청취 실험을 통해 이루어졌다. 10명의 피실험자에게 가라사대[8]와 본 피치규칙을 적용한 합성 시스템의 문장들을 들려주고 자연성을 비교 평가하도록 하였다. 평가 단계는 본 합성 시스템이 가라사대와 비교해 1) 매우 어색하다. 2) 어색하다. 3) 비슷하다. 4) 자연스럽다. 5) 매우 자연스럽다. 와 같이 5단계로 구분하였다.

청취 실험 결과는 그림 1과 같다.

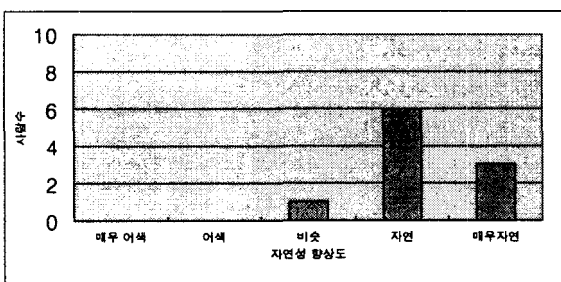


그림 1. 지속시간의 자연성 향상도

그림 1에서 보면 10명의 피실험자 가운데 1명이 '비슷하다', 6명이 '자연스럽다', 3명이 '매우 자연스럽다'에 응답하여, 최고점을 5로 할 때 전체 평균 4.2이다. 이것은 본 지속시간 규칙이 기존 가라사대에 비해 합성음의 자연성을 향상시켰다는 것을 보여준다.

5. 결론

지속시간 변화법칙을 구하기 위해 표준말을 사용하는 20대 중반의 여자 10명을 피실험자로 삼아, 457개의 문장을 하루에 1번씩 3일에 걸쳐 3회 발화시켰다.

그 결과 자, 모음의 고유 지속시간을 구했고, 자, 모음이 단어 내에서 차지하는 위치(단어 초, 단어중간, 단어말)에 따라 지속시간이 변화한다는 것, 한 어절 내 음절 수 증가에 따라 한 음절 당 지속시간이 감소한다는 것, 인접 음소에 따라 자음, 모음의 지속시간이 변화한다는 것, 문장 유형(평서문, 의문문, 명령문, 인용문) 중 예/아니오 의문문에서 마지막 음절의 지속시간이 가장 짧다는 것, 구나 절의 경계에서 마지막 음절의 지속시간이 길어지고, 휴지부가 존재한다는 것, 발화 속도 증가에 따라 문장 전체 길이가 증가하는데 휴지부의 길이가 큰 폭으로 변화한다는 것, 경, 격음 앞에서의 묵음의 길이 등에 관한 법칙을 구했다.

우리말의 지속시간에 관한 연구는 그리 많지 않고, 특히 실험적 연구는 그 수가 매우 적다. 본 연구는 여성 음의 지속시간 변화법칙을 다루었으나, 남성 음에 대한 연구도 추가로 실시하여 남성 음과 여성 음의 지속시간 변화법칙을 비교하는 것도 필요하다. 이것은 앞으로의 연구과제로 남겨 두겠다.

참고 문헌

1. 고 도홍, Declarative intonation in Korean : an acoustical study of F0 declination, 한신 출판사, 1988.
2. 구 회산, An experimental acoustic study of the phonetics of intonation in Korean, 한신 출판사, 1988.
3. 성 철재, 한국어 리듬의 실험음성학적 연구 - 시간 구조와 관련하여, 서울대학교 박사학위 논문, 1995.
4. 김 진영, 성 평모, "한국어의 억양에 관한 연구", Korea-Japan joint symposium on acoustics, pp. 292-297, 1991.
5. 임 운천, 안 수길, "독립 단어에서의 운율에 관한 연구", Korea-Japan joint symposium on acoustics, pp. 280-291, 1991.
6. 지 민제, 최 운천, 김 상훈, "우리말 소리의 길이 : 실험음성학적 연구," 제5회 한글 및 한국어 정보처리 학술발표 논문집, pp. 119-130, 1993.
7. 성 유나, 이 양희, "한국어 음운 지속시간 모델화를 위한 특징 분석," 제 13회 음성통신 및 신호처리 워크샵 논문집, pp. 301-306, 1996.
8. 권 철홍, 정 원국, 구 준모, 김 형순, "한국어 문자음성 변환 시스템 : 가라사대," 한국통신학회지 제 11권 9호, pp. 17-25, 1994.