

국어 숫자음의 음절구조에 대한 음향적 분석

양병곤
부산대학교 영어교육과

An Acoustical Study on the Syllable Structures of Korean Numeric Sounds

Byunggon Yang
English Education Dept., Pusan National University
E-mail : bgyang@pusan.ac.kr

Abstract

The purpose of this study was to examine the syllable structures of ten Korean numeric sounds produced by ten subjects of the same age. Each sound was normalized and divided into onset, vowel, and coda sections. Then, acoustical measurements of each syllable were done to compare the ten sounds. Results showed that there was not much deviation from the grand average duration and intensity for the majority of the sounds except the two diphthongal sounds on which their boundary points varied among the speakers. Some syllable boundaries were quite obvious while others were ambiguous. There seemed some tradeoff among the syllable components depending on their acoustic features.

I. 서론

음절은 음성분석과 합성에 중요한 역할을 한다. 음절은 누구나 쉽게 몇 개인지 확인해 줄 수 있지만, 조음적으로나 음향적으로 정의하기는 어렵다(Sloat, Taylor & Hoard, 1978; Gigerich, 1992; Roach, 1999). 물론 화자마다 다르게 발음하는 방언적인 차이도 있을 수 있지만, 그 차이는 한두 개에 불과하다. 음

절은 초성, 중성, 종성으로 나뉘어진다(McMahon, 2002). 초성과 종성은 주로 자음으로 되어 있고, 음절에서도 가장 또렷한 중성에는 모음이나 공명자음이 주를 이룬다. 이러한 구조를 음향적으로 분석하는 데는 초성과 중성 또는 중성과 종성 사이의 경계선이 모호하여 음소수준에서 분리를 하려면 매우 어렵다(Port, Al-Ani & Maeda, 1980; Port, Dalby & O'Dell, 1987).

기존 연구 가운데 Yang(2001)은 보다 안정된 숫자음을 이용하여 화자확인을 연구하기 위해 9개의 숫자음의 음향적 특징을 분석해 재합성한 뒤 원래의 소리와 연이어 피험자에게 들려주면서 같은 숫자음으로 들리는 음향적 변수의 범위를 구하였다. 음향적 측정값은 매우 광범위하게 변하기 때문에 사람이 듣는 단위의 범위를 찾기 위한 시도였다. 그 결과 강도값에서는 6.6 dB, 피치값에서는 10.5 Hz, 첫 번째 세 개의 포먼트 값에서는 5.9%의 변화를 주어도 동일한 숫자음으로 받아들임을 보고했다. Yang & Kang(2002)의 연구에서는 화자확인을 하기 위한 방법으로 숫자음의 전체 지속시간을 100등분하여 각 지점마다 스펙트럼 에너지를 구한 뒤 다섯 개의 대역으로 여과하여 각 영역의 에너지 합을 구한 뒤 화자내와 화자간의 숫자음 별로 비교하여 상관계수와 차이값을 구했다. 비록 안정적인 화자확인용 변수로 확정할 수 없었지만, 이러한 스펙트럼 에너지 정보는 음절의 모음과 자음 부분을 구분하는데 좋은 도구가 될 것으로 생각된다. 영어음절 구분에 대한 연구로 Yang(2006)에서는 자음+a를 연속적으로 발성한 원어민의 목소리를 프라트 편집창을 통해

확대하여 보면서 몇 가지 시각적인 경계선 기준을 적용하여 보다 객관적인 측정값을 구하려 시도했다. 예를 들어, 초성과 중성간의 경계선을 VOT를 구할 때 사용하는 지점인 첫 번째 상대진동펄스를 기준으로 가장 가까운 영점교차점을 찾았고, 측음과 모음은 모두 공명음으로 애매한 경우가 많았는데, 그 사이의 경계를 각 성대 펄스의 변화를 관찰하여 모음에서 보이는 세 개씩의 정점이 보이는 구간이 두 개로 바뀐 지점을 찾았다. 또 비음과 모음의 결합에서도 이런 방법을 이용하였다. 하지만 이러한 방법에서도 여전히 경계선을 구하는데 어려움이 있어, 일관성 있게 적용하는 방법을 찾아보게 되었다.

이 연구에서는 10명의 화자가 발성한 국어 숫자음의 음절 경계를 좀 더 객관적인 기준에 따라 찾기 위해, 프라트의 스크립트를 사용하여 기존 연구의 스펙트럼 강도의 누적값을 구한 다음 이 값들의 변화가 일어나는 경계점을 찾아 음절을 구분하고 음절 구성 요소 간에 어떤 통계적 특징이 있는지 살펴보고, 연구과정에서 또 다른 문제점은 어떤 것이 나타나는지 알아보았다.

II. 연구방법

분석에 사용한 숫자음은 SiTEC에서 수집한 음성자료로 10명의 같은 나이의 남성 화자가 발음한 것을 이용했다. 음성자료는 방음실에서 Senheizer HMD224X 마이크로 16비트 16 kHz의 표본속도로 녹음되었다. 사용된 자료를 발성한 화자는 모두 21세로 서울에서 태어나 생활한 사람들이다. 음성자료는 모두 10개의 숫자음이며, 단모음으로만 된 “이, 오”와, 자음과 모음으로 된 “사, 구”, 모음과 자음으로 된 “일”, 자음+모음+자음의 형태로 된 “영, 삼, 육, 칠, 팔, 십”으로 나눌 수 있는데, 실제 발음에서는 “육”과 “십”의 중성을 파열음으로 발음한 경우가 적었고, 이중모음의 첫 음인 전이음과 모음의 경계가 애매했다. 숫자음 “영”에서도 마찬가지로 전이음의 경계가 뚜렷하지 않고, 중성과 중성의 비음경계가 애매했다. 음절분리과정은 프라트 스크립트를 이용해 먼저 각 숫자음의 에너지를 정규화시켰고 초성 앞부분과 중성 뒷부분에 각각 100 ms의 묵음구간을 두어 주변의 강도값의 영향을 받지 않도록 했다. 초성과 중성, 중성과 중성 사이의 경계지점은 프라트 스크립트를 이용하여 시각적으로 먼저 경계선을 찾아 각 지점의 중심점에 마우스를 두어 일정한 간격으로 스펙트럼에너지를 구하고, 중성의 기준값에서 한 표준편차의 범위를 갑자기 벗어나는 시점을 화면에 프린트하여 이 값이 적절한지 판단하여 경계점으로 받아

들이면 가장 가까운 상대진동펄스점을 찾아 시간점을 표시하거나 문제가 있을 경우 확대한 후 다른 지점을 찾기도 하였다. 이렇게 각각의 숫자음을 초중중성으로 분리한 다음 각 구간의 지속시간비율이나 누적강도값의 비율을 구하여 살펴보았다.

III. 결과 및 논의

모든 숫자음의 평균 지속시간은 332 ms이고 강도값은 80 dB로 나타났다. 숫자음 “이”는 평균 342 ms로 발음되어 숫자음 “오”보다 13 ms정도 더 길게 발음했고, 평균 강도값은 81 dB를 보였으며, 100개 지점의 강도 변화량은 비슷한 모양을 보였다. 숫자음 “사”와 “구”는 약 387 ms로 발음되었으며, 숫자음 “십”이 286 ms로 짧게 나타났는데, 이는 숫자음 중 가장 짧은 “육”의 235 ms와 함께, 중성의 묵음구간이 포함되지 않았기 때문으로 여겨진다. 만약 묵음구간도 포함하면 다른 숫자음과 비슷한 길이의 발음으로 될 것으로 여겨진다. 다시 말해서, 국어 숫자음의 전체길이에 따라 초성과 중성, 중성이 상호작용하여 적절한 비율을 차지하는 것으로 생각된다. 그림 1은 숫자음 “사, 구, 십”의 100개 지점의 평균값을 구해 나타내어보았다.

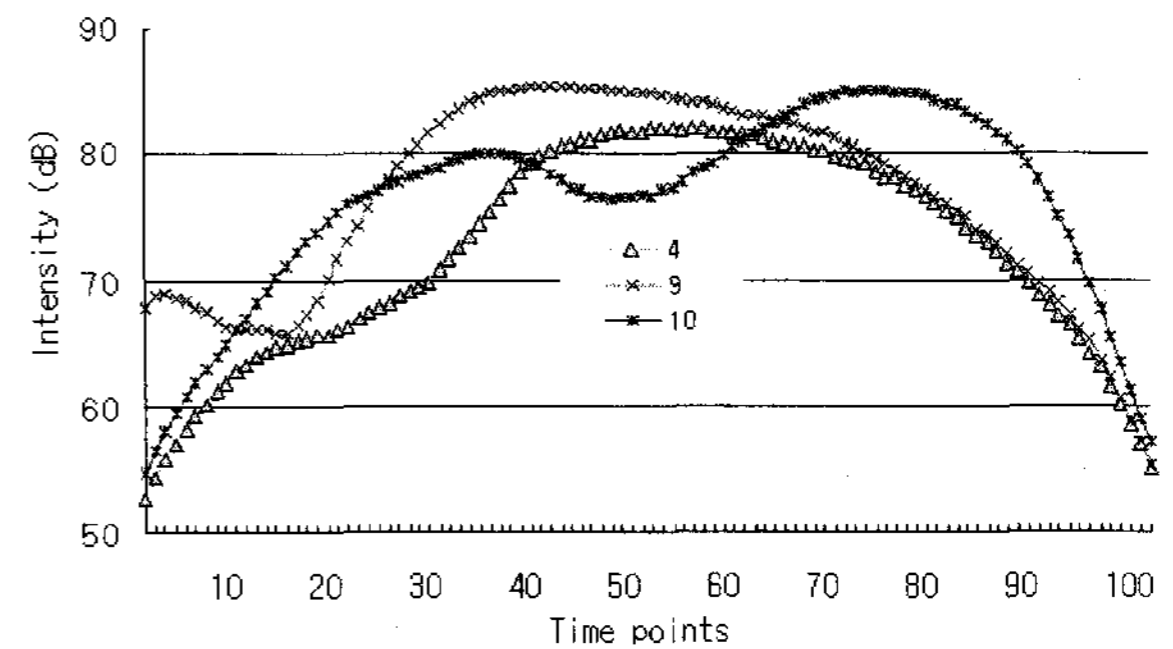


그림 1. 숫자음 “사, 구, 십”의 평균강도곡선

그림에서 보면 초성에서 초성과 중성의 경계점을 보이는 움푹 들어간 부분까지의 시간이 “구”에서 가장 짧고 “사” 다음에 “십”으로 이어진다. 물론 여기서 100 등분한 시간점이기 때문에 숫자음 “십”의 길이는 상대적으로 길게 나타났지만, 실제로는 “사”와 약 40 ms의 차이를 보였다. “십”의 전체 음절에서 중성이 차지하는 비율은 약 45%로 절반이 채 안되었다.

숫자음 “일”에서는 음절 경계선이 다소 명확하여 지속시간 측정편차가 적은 반면 숫자음 “영”에서 점점 높아져, 숫자음 “육”에서는 모음 편차가 146 ms나 보였다. 이는 이중모음의 계속되는 변화에서 비록 스펙트럼 강도누적합의 경계선을 이용하더라도 여전히 일관된 지점을 찾기가 어려우며 화자마다 다른 발음 습

관이 측정값의 변화량을 높인 것으로 나타났다. 숫자음 “일”에서는 중성구간의 시작이 음절의 1/3지점에서 시작되었고, 중성과 중성간의 경계점에서는 일시적으로 강도가 낮아지는 흐름을 보이는 화자가 많았다. 비음이 포함된 숫자음 “영”에서는 중성의 누적강도값이 반공명작용으로 낮게 나타났다(Picket, 1987:77, 124).

그림 2는 숫자음 “삼, 칠, 팔”의 평균강도곡선을 보인다.

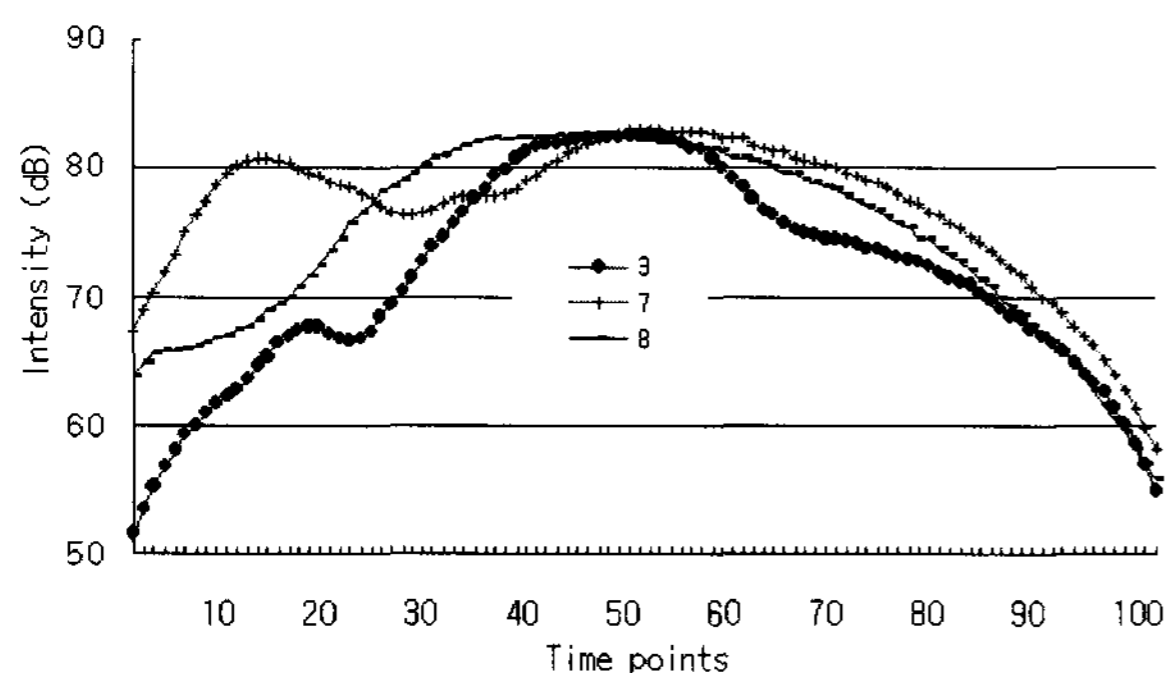


그림 2. 숫자음 “삼, 칠, 팔”의 평균강도곡선

세 숫자음의 평균 지속시간은 334 ms에서 391 ms의 범위에 있으며, 스펙트럼강도의 누적값은 비음이 포함된 “삼”이 가장 낮게 나타났다. 지속시간의 측정에서는 편차가 다소 낮았는데, 숫자음 “삼”의 중성과 중성간의 경계가 매우 뚜렷하기 때문이다. 숫자음 “삼”에서는 초성과 중성과 중성의 각각 지속시간 비율이 31%, 30%, 39%를 차지하여 매우 이상적인 분포를 보인다. “칠”에서는 33%, 24%, 40%로 중성부분이 다소 짧은 편이며, “팔”에서는 19%, 30%, 51%로 초성에서 다소 짧아진 시간을 중성이 가져간 분포를 보이고 있다. 앞서 보았던 “사”에서는 초성과 중성비율이 31:69%이고, “십”에서는 55:45%, “구”에서는 22:78%를 보였다. 중성이 유성음인 “영”에서는 51:49%, “일”에서는 36:64%를 보였고, 숫자음 “육”에서는 63:37%를 보였다. 이러한 비율은 스펙트럼강도의 누적값에서도 거의 비슷한 값으로 나타났다. 따라서 음절내에서 각 구성요소 간 시간 배분은 해당 자모음의 특성이 어느 정도 반영되어 나타남을 알 수 있다. 숫자음 “팔”의 중성과 중성 간 경계선도 앞에서 보았던 숫자음 “일”과 같이 매우 독특한 에너지 함몰을 보이기 때문에 이를 기준으로 경계선을 찾아도 될 것이다. “일”의 중성은 화자마다 음절 내에서 상대적으로 높은 강도로 발음하였다.

IV. 결론

본 논문에서는 음성합성과 분석에 흔히 사용되는 열개의 국어 숫자음에 대한 음절구조를 분석하기 위해 열 명의 화자가 발음한 숫자음을 대상으로 전체지속시간을 100 등분하여 각 지점마다의 스펙트럼강도의 누적값을 구한 다음 이 값들의 변화지점을 음절경계선을 찾는데 이용하여 초성, 중성, 종성으로 나누었다. 이어서, 각 구성요소별 음향적 측정값의 비율과 분포를 살펴보았다.

그 결과 전체 숫자음의 발음시간은 332 ms로 발음되었고 평균 강도는 80 dB이었으며, 이중모음이 들어간 숫자음에서는 음절 경계가 뚜렷하지 않아 상대적으로 높은 측정 편차를 보였으며, 자음+모음의 구조로 된 일부 숫자음들은 매우 뚜렷한 경계점을 보여서 차후에 자동으로 분리하는 방법을 사용할 수 있음을 관찰할 수 있었다. 덧붙여, 음절 내에서는 전체 구성성분이 차지하는 비율이 대체로 비슷하게 나타났지만, 각 자모음의 특성에 따라 한 부분이 짧아지면 다른 부분이 길어지는 방식으로 상호 작용을 함을 알 수 있었다. 이러한 음절구조분석에 대한 보다 일관성 있고 객관적인 기준이 제시된다면, 국어 음절구조에 대한 보다 일반적인 결론을 제시할 수 있을 것으로 기대된다.

참고문헌

- Gigerich, H. J. 1992. *English Phonology: An Introduction*. Cambridge: Cambridge University Press.
- Roach, P. 1999. *English Phonetics and Phonology*. Cambridge: Cambridge University Press.
- Picket, J. M. 1987. *The Sounds of Speech Communication: A Primer of Acoustic phonetics and Speech Perception*. Austin, Texas: Pro-ed.
- Port, R. F., Al-Ani, S. & Maeda, S. 1980. “Temporal compensation and universal phonetics.” *Phonetica* 37, 232-252.
- Port, R. F., Dalby, J. & O'Dell, M. 1987. “Evidence for mora-timing in Japanese.” *Journal of the Acoustical Society of America* 81, 1574-1585.
- Sloat, C., Taylor, S. H. & Hoard, J. E. 1978. *An Introduction to English Phonology*. Englewood Cliffs, N.J.: Prentice-Hall International, Inc.
- Yang, B. 2001. “Speaker variation in number production by males.” *Speech Sciences* 8(3), 93-104.
- Yang, B. & Kang, S. 2002. “A study on speaker identification by difference sum and correlation coefficient of intensity levels from band-pass filtered sounds.” *Speech Sciences* 9(3), 3-16.
- Yang, B. 2006. “An acoustical study of English CV syllables.” *Speech Sciences* 13(4), 127-140.