한국어 음성합성에서 음운지속시간 모델화

이양희(동덕여대 전자계산학과)

<차례>

1. 서론
2. 음운지속시간 제어요소의 분석
2.1 음성데이터
2.2 음운지속시간의 특징
2.2.1 문절내 음절수의 효과
2.2.2 문절내 음절위치의 효과
2.2.3 인접 음운의 효과
3. 음운지속시간의 모델화
3.1 제어요소 및 모델
3.2 예측결과
4. 결론

<요약>

Segmental duration modelling for Korean text-to-speech synthesis

YangHee, Lee(Dongduck Women's Univ.)

본 논문에서는 자연스러운 음성을 합성하기 위하여, 한국어 음운지속시간의 변화에 있어서 문절과 구내의 음절수와 음절의 위치에 의한 영향과 인접하는 음운의 영향에 대하여 통계적으로 분석하였고, 분석된 시간 특징을 제어 요소로 하는 회귀트리를 생성하여 음운 지속시간을 모델화하였다. 또한, 제안된 음운 지속시간 모델에 의해 예측실험을 행하여, 측정치와 예측치간의 다중 상관계수가 0.74 정도이고, 각 음운의 예측오차의 75%이상이 25ms이내로 제안된 모델의 타당성이 입증되었다.

1. 서론

자연스러운 리듬과 템포를 갖는 음성을 합성하기 위해, 정교한 음운지속시간의 모델화가 필요하다. 많은 언어들에 대해 음운지속시간의 모델화가 연구되고 있다[1][3][5][7].

음성데이터로 부터 음운 지속시간 제어규칙을 자동적으로 생성하기 위해, 최근 통계적인 방법들이 이용되고 있다. 선형회기모델[3], 회귀 트리모델[4], 적의 합(Sum of product)모델[5]등이 규칙의 최적화 및 규칙을 자동생성하기 위해 제안되었다.

음절의 장단에 의해 의미를 변별하는 한국어에 있어서 운율의 다른 요소보다도 음운의 지속시간 모델은 특히 중요하지만, 자연음성에 대한 시간특성에 대한 분석 및 모델화에 대한 연구가 별로 되어 있지 않다. 최근 한국어의 모음에 있어서 전후 인접하는 2개의 자음의 효과를 조사하여, Klatt모델에 적용한 결과가 보고되었으나[6], 전후 각각의 음소의 영향에 대하여는 조사되지는 않았다.

또한, 일반적인 규칙을 자동생성하기 위해 통계적인 모델을 필요로 하고 있으나 이에 대한 연구가 거의 행하여져 있지 않은 상태이다. 따라서 본 논문에서는 정확한 음운지속시간을 모델화하기 위해 2절에서는 한국어의 음운 지속시간의 변화에 있어서, 문절과 구내의 음절 수와 음절의 위치에 의한 효과, 그리고 전후음운 각각에 의한 효과를 통계적으로 분석하고, 3절에서는 이 요소들을 이용하여 회귀트리로 음운지속시간을 모델화한다.

2. 음운지속시간 제어요소의 분석

2.1 음성데이터

회귀트리에 의한 음운지속시간을 모델화하기 위해서는 그 모델에서 사용되는 특징요소의 선택이 매우 중요하다. 이 절에서는 한국어의 시간특징을 통계적으로 분석한다.

시간특징들을 분석하기 위해서 이용된 음성데이터는 7명의 화자(대학생, 남:3명, 여:4명)가 3종류의 템포(빠르게, 보통, 느리게)로 낭독한 336개의 발화(문)이다. 이 음성데이터에 포함된 음절의 수는 635개이고, 음소의 수는 1441개이다.

2.2 음운지속시간의 특징

음절 및 음운지속시간의 특징 분석에 있어서 각 음절 유형 및 음소의 고유 지속시간의 영향을 배제하기 위하여, 각 음절 유형 및 음소에 대해 Z Score로 지속시간을 정규화한다. 정규화 지속시간은 다음과 같은 식으로 구한다.

$$z_{ip} = (x_{ip} - \overline{x_p}) \div \sigma \qquad (1)$$

x_{ip} : 음소 p의 지속시간에 대한 i번째의 관측치

$\overline{x_p}$: 음소 p의 지속시간에 대한 평균치

σ : 음소 p의 지속시간에 대한 표준편차

2.2.1 문절내 음절수의 효과

이 절에서는 CV와 CVC음절유형내 각각의 음운의 Z Score에 의한 정규화 지속시간을 발화 템포별로 구하여, 문절내 음절의 수에 따라 자음 및 모음의 지속시간이 어떻게 변화하는가를 조사한다. 음절수에 따른 음운의 지속시간 변화는 그림1과 같다.

그림 1은 CV음절에 있어서 문절내 음절수에 의한 자음의 지속시간 변화를 보통 템포에서 나타내고 있다.

발화 보통템포에 있어서, 자음의 지속시간은 문절내 음절수의 증가에 따라 서서히 감소하지만, 감소폭이 작다. 빠른 템포의 경우 음운의 고유 지속시간이 있으므로 음절수의 증가에 의한 감소 폭이 다른 템포보다도 작다. 문절내 음절수에 의한 CV음절의 모음 지속시간 변화는 그림 2와 같다.

모음의 지속시간 변화는 자음과 비슷하지만, 변화의 폭이 자음의 변화폭보다 다소 크다. 또한, CVC음절에 대한 자음의 지속시간 변화로 CV음절과 유사하다.

CV, CVC음절유형 모두 음절구성요소의 지속시간의 길이는 음절수의 증가에 따라 감소하지만, 탄성이 큰 모음의 변화 폭이 탄성이 작은 자음의 변화보다 크고, CVC음절의 변화폭이 CV음절의 변화폭보다 크다. 또한 음운의 고유지속시간 때문에 빠른템포의 변화는 다른 템포의 변화폭 보다 작다. 이것은 조음하기 위해, 최소한의 고유지속시간이 필요하기 때문에, 음운 지속시간이 줄어드는 폭이 길어지는 폭보다 작다는 것을 나타낸다.

2.2.2 문절내 음절위치의 효과

각각의 음절 유형에 있어서, 음절을 구성하는 음운에 대하여, 그 음운의 영향을 배제하는 정규화 지속시간으로 문절내 음절위치의 효과를 조사 한다.

CV형 음절에 있어서, 문절내 위치에 의한 각 음운의 정규화 지속시간은 그림3.과 4와 같다.

이들 그림으로 부터, 모음의 경우 처음과 중간 음절에서는 전반적으로 짧아지고, 마지막 음절에서는 길어지는 경향이 있음을 알 수 있다. 그러나, 자음경우 처음 음절에서는 길어지는 경향이 있음을 알 수 있다. 즉, 지속시간의 신장폭의 크기는

마찰음/s/ > 비음 > 유기음, 경음, 무성자음/g,d,b,z,h/과 같고, 중간 음절에서는 유기음과 경음경우 평균길이를 갖지만 다른 자음들은 짧아진다. 특히, 자음/g,d,b,z/는 유성음화 때문에 다른 자음 보다 더욱 짧아지는 경향이 있다. 즉, 변화폭의 크기는 유성화 자음/g,d,b,z,h/ > /s/, 비음, 유음 > 유기음, 경음과 같다. CVC음절에 대해서도 유사하지만 모음의 길이는 크게 변하지 않는다.

2.2.3 인접 음운의 효과

한국어에 있어서 모음의 전후 양쪽 음운의 조합에 의해 모음 지속시간 변화를 조사하여 타이밍제어에 이용한 연구가 문헌[5]에 보고 되었다. 그러나 정확한 음운의 특징을 얻기 위하여 앞의 음운, 뒤의 음운으로 분리하여 각각의 효과를 조사할 필요가 있다.

이 절에서는 CV음절과 CVC음절에 있어서 각 음운의 앞의 음운과 뒤 음운의 영향을 분리하여 조사한다. CV음절에 있어서 모음지속시간에 대한 앞 자음의 효과는 그림 5와 같다. 이 그림으로부터 모음지속시간은 앞의 자음에 의해 현저하게 변한다. 자음 종류에 의한 모음의 길이는 다음과 같음을 알 수 있다.

비음, 유음 > 유성 파열음 > 유기음, 경음, 마찰음, 파찰음

또한, CV음절에 있어서, 모음지속시간에 대한 뒷자음의 효과는 그림 6과 같다. 여기에서 모든 뒷 자음에 의해 모음의 지속시간이 전반적으로 짧아지는 경향이 있다. 특별한 자음에 의해 모음의 지속시간이 크게 변화하지 않는다는 것이다. 이로부터 CV음절에 있어서, 모음의 지속시간은 음절프레임에 의해 CV음절의 자음과 모음간의 조음결합이 다른 음절의 자음과 조음결합이 강하기 때문에 앞의 자음의 영향이 뒷 자음의 영향보다 크다고 생각된다.

그림 5에서 앞의 자음이 길어지면, 모음이 짧아지고 자음이 짧아 지면 모음

이 길어지는 경향이 보인다. 이것은 음절의 고유지속시간이 있어서 음절내에서의 음운의 상호보상 때문이라고 생각된다. CVC음절에 있어서 모음의 앞 자음의 지속시간길이에 대한 그 뒤 모음의 정규화 지속시간의 변화와 그림 6에 뒷 자음의 지속시간길이에 대한 그 앞 모음의 정규화 지속시간의 변화는 그림 6과 같다. 모음지속시간 길이에 대한 앞의 자음의 영향은 CV음절과 유사하지만, 뒷 자음 즉 종성의 영향은 종성의 음운과는 관계없이 일정하다. 이것으로 부터, 모음의 지속시간 길이는 뒤의 음운보다 앞의 음운의 영향이 큰 것을 알 수 있다. 또한 자음 지속시간 길이에 대한 앞 음운의 효과는 그림7과 같다.

앞의 음운이 모음인 경우, 자음의 지속시간 길이의 변화는 거의 없지만, 앞의 음운이 종성의 경우, 비음인 종성 뒤의 자음이 짧아 진다. 이것은 자음의 유성음화가 강하게 나타난 것을 의미한다.

3. 음운지속시간의 모델화

3.1 제어요소 및 모델

관측데이터로 부터 일반적인 제어규칙을 추측하기 위한 통계모델로서, 설명변수 공간을 축차 트리로 분할하는 것으로 카테고리간 의존관계등에 대한 분포의 비선형성을 표현할 수 있는 회귀트리를 이용한다. 이와 같은 통계적방법에서는 특징집합의 선택이 매우 중요하다. 본 논문에서 사용된 특징집합은 앞절에서 분석된 음운지속시간을 변화시키는 중요한 요인들이다. 따라서 이들 요인들을 이용하여 음운지속시간을 모델화 한다. 음운지속시간을 예측하기 위한 중요한 요인은 다음과 같다.

- 음절유형 : 예측할 음운을 포함하는 음절유형
- 음운 콘텍스트 : 예측할 음운, 예측할 음운의 좌우 2개음운
- 문절내 위치 : 문절의 처음과 마지막 및 중간 음절, 문절내 음절수
- 구내 위치 : 구의 처음과 마지막 및 중간 음절, 구내 음절수

여기에서, 음절유형 = {V, SV, SVC, CV, CSV, VC, SVC, CVC, CSVC}인 9종류로 분류되고, 각 음운은 조음양식과 조음위치에 의해 자음은 13종류 = {파열연음, 파열기음, 양순비음, 치경비음, 양순비음과 치경비음 받침, 유음 받침, 연구개비음 받침, 무성음 받침, 치경마찰음, 후두마찰음, 파열경음, 마찰과 파찰 경음, 반모음}로, 모음은 4종류 = {저, 중, 고, 복모음}로 분류된다. 문절과 구내의 음절의 위치는 3종류 = {시작, 중간, 마지막}으로 분류되고, 문절내

음절수는 7종류 = {1, …, 7}로 분류되고, 구내 음절수는 16종류 = {1, …, 16}로 분류된다.

3.2 예측결과

7명의 화자에 의한 보통 템포의 16개 발화(문)중 15개 발화(문)가 트리를 생성하기 위한 훈련데이터로 사용되었고, 나머지 1개의 문이 각 템포로 3개의 발화는 테스트 데이터로 사용되었다. 모든 16개 문에 대해 동일하게 예측이 행해졌다. 그 결과는 그림 8과 같다.

이 결과에 의하면, 화자에 따라 다르지만, 여성쪽이 남성보다 더 좋게 예측되었다. 또한, 빠른 템포보다 느린 템포의 발성이 더 양호하게 예측되었다. 평균 다중상관 계수는 표1.과 같다.

표1. 각 템포별 예측결과

tempo	normal	fast	slow
평균 다중상관	0.74	0.69	0.74

음성데이터내의 모든 음운의 지속시간 분포는 그림 9와 같다.

이 그림으로부터 자음과 모음의 75%이상이 25ms이내의 오차로 예측할 수 있음을 나타낸다. 따라서 여기에서 생성된 음운지속시간 예측모델이 타당함을 나타낸다. 이 모델에 품사정보등 문법적인 정보를 부가함에 의해 더욱 양호한 예측이 가능하다.

4. 결론

한국어 음성합성에 있어서 자연스러운 음성을 합성하기 위하여, 한국어의 타이밍특징을 통계적으로 분석하였다. 그 결과, 1) CVC음절에서는 문절 또는 구의 시작음절이 길어지고, CV음절에서는 마지막 음절이 길어진다 2) 음절과 음운의 지속시간은 문절 또는 구내의 음절수에 반비례한다 3) 모음의 지속시간은 다음에오는 자음보다도 앞에 오는 자음의 영향을 크게 받는다는 것을 분명히 하였다. 또한 이들 특징집합을 회귀트리에 이용하여 음운지속시간을 모델화하였고, 제안된 모델을 사용하여 예측한 음운지속시간을 평가하여, 그 모델의 유효성을 확인하였다. 금후과제로는 제안된 모델의 제어요소에 품사정보 및 문법적 정보를 부가하여 모델을 개선하고자 한다.

<참고문헌>

[1] D.H.Klatt, "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," Journal of the Acoustical Society of America, 59, 5, pp1208-1221, 1976.

[2] H.B.Lee et al., "An experimental phonetic study of speech rhythm in standard Korean," Proc. of ICSLP, pp1091-1094, 1994.

[3] N. Kaiki, K. Takeda and Y.Sagisaka, "Linguistic properties in the control of segmental duration for synthesis," Talking machines : Theories, Models, Designs, pp255-263, 1992.

[4] M.D.Riley, "Tree-based modelling of segmental duration," Talking machines : Theories, Models, Designs, pp265-273, 1992.

[5] Jan P.H. van Santen, "Deriving text-to-speech duration from natural speech," Talking machines : Theories, Models, Designs, pp275-285, 1992.

[6] S.H.Kim and J.C.Lee, "Korean text-to-speech system using time domain-pitch synchronous overlap and add method," International conf. on SST, vol.2 pp315-320, 1994.

[7] W.N.Campbell, "Syllable based segmental duration," Talking machines : Theories, Models, Designs, pp211-223, 1992.

[8] Y.Sagisaka, "日本語音聲の韻律的特徵とその計算モデル," 日本音響學會秋季講演論文集, 1992, pp295-298.

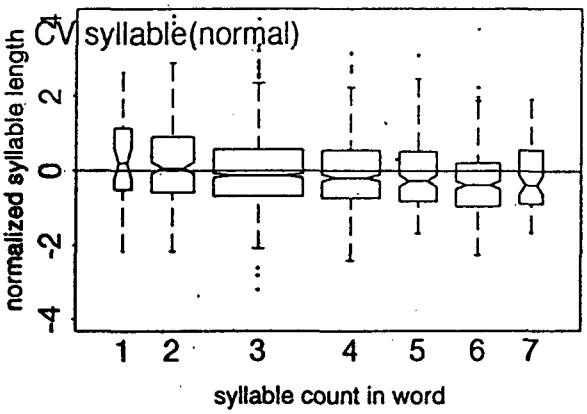

그림1. 문절내 음절수에 의한 자음의 지속시간 변화 (CV음절)

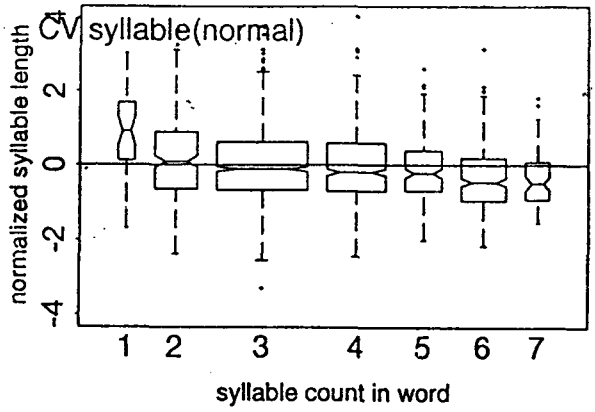

그림2. 문절내 음절수에 의한 모음의 지속시간 변화 (CV음절)

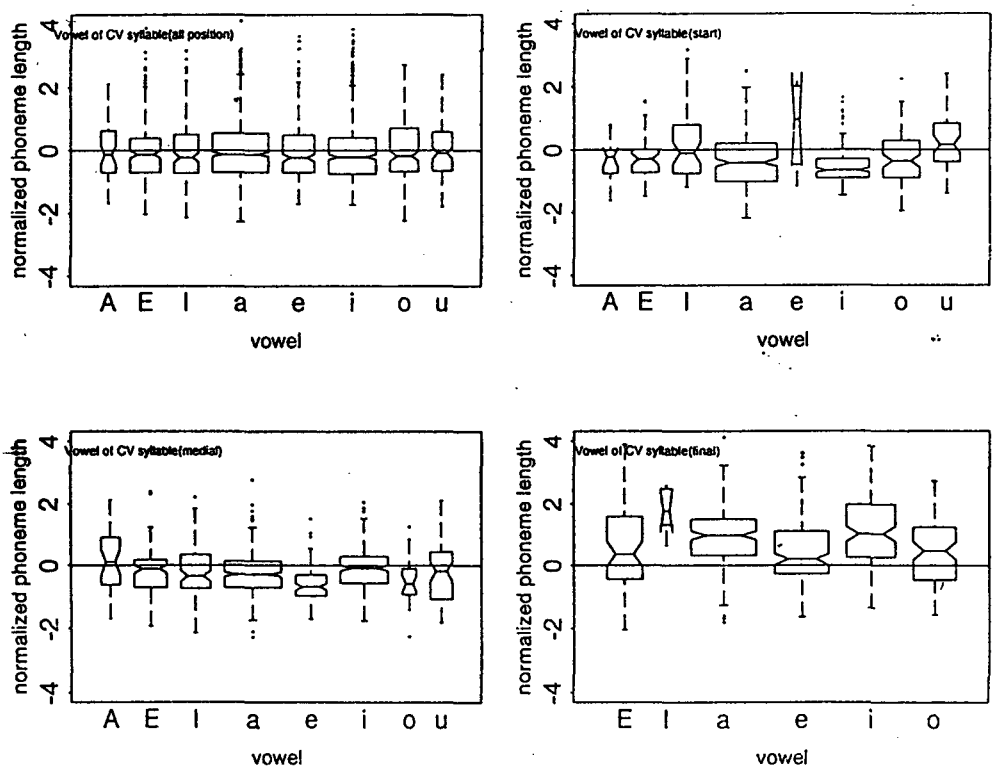

그림3. 모음 지속시간 변화에 대한 음절위치의 영향 (CV음절)

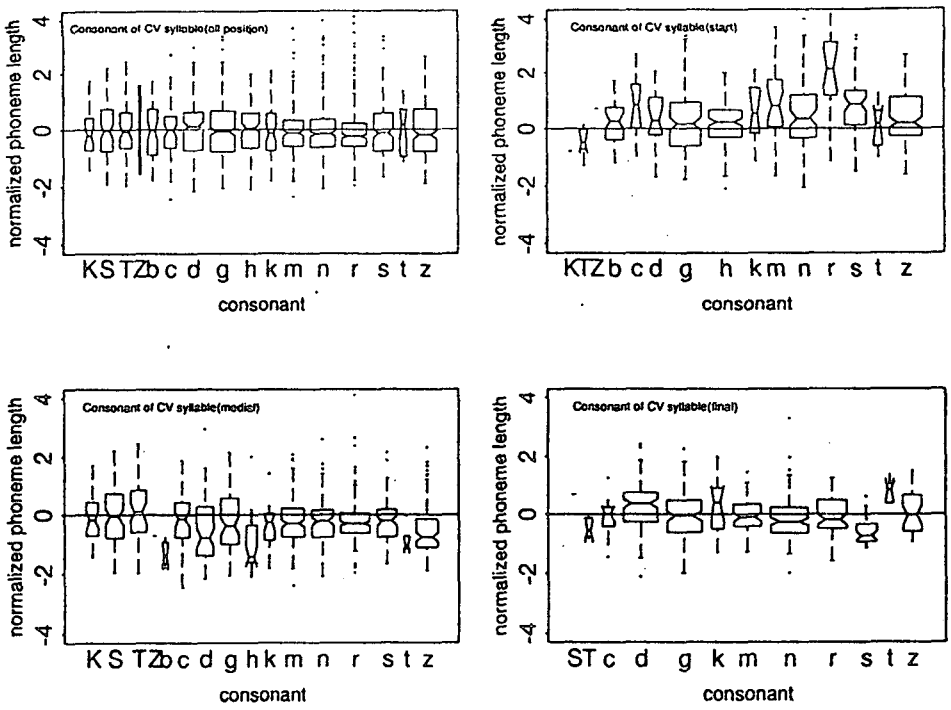

그림4. 자음 지속시간 변화에 대한 음절위치의 영향 (CV음절)

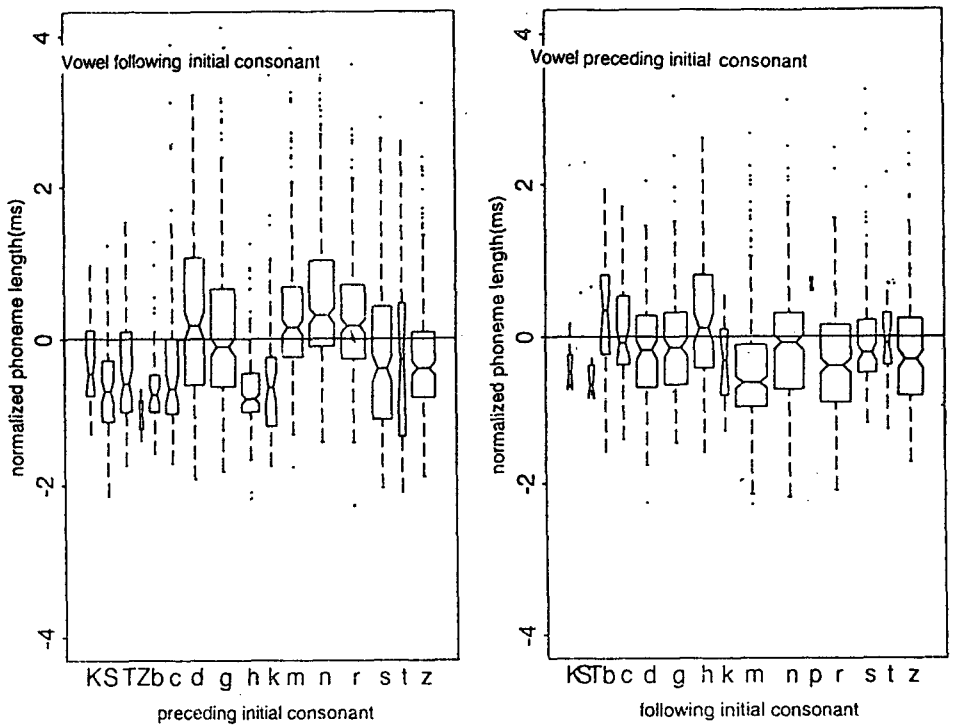

그림5. 모음 지속시간 변화에 대한 인접자음의 효과(CV)

(좌) 앞의 음=초성

(우) 뒤의 음 = 다음 음절의 초성

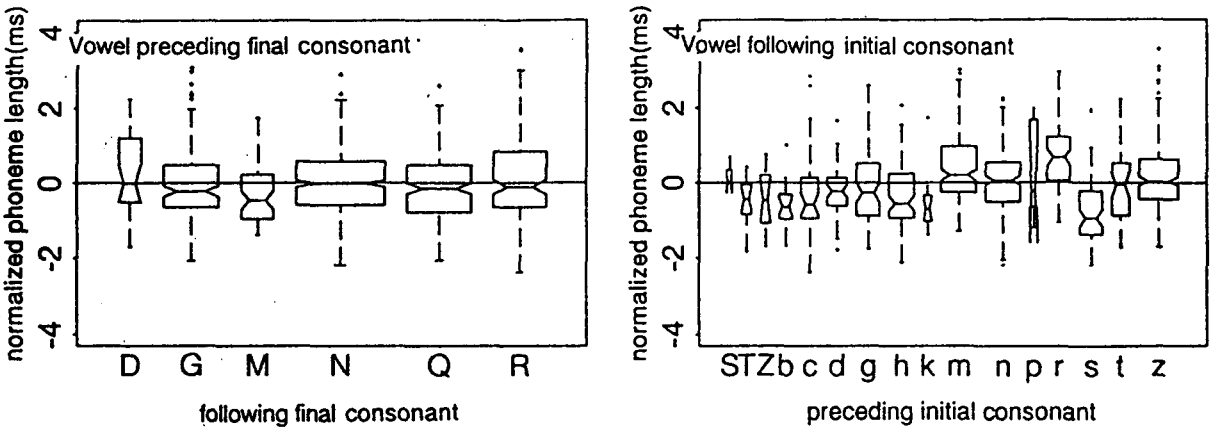

그림6. 모음 지속시간 변화에 대한 인접자음의 효과(CVC)
(좌) 인접자음 종성 (우) 인접자음 종성

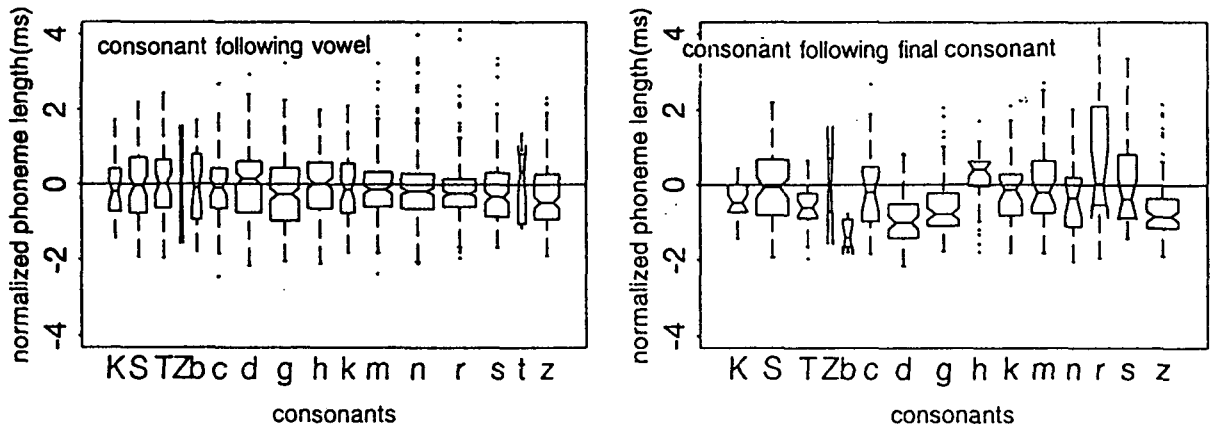

그림7. 자음 지속시간 변화에 대한 선행음의 효과
(좌) 앞의 음 = 모음 (우) 앞의 음 = 종성

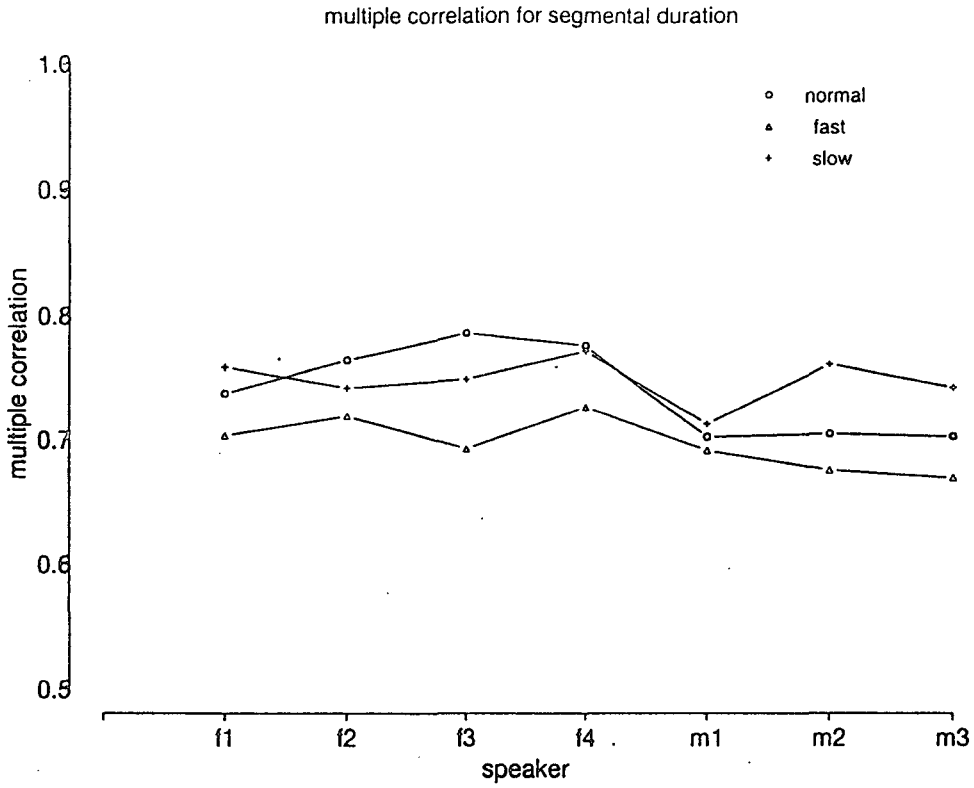

그림8. 예측과 관측 지속시간 사이의 다중상관

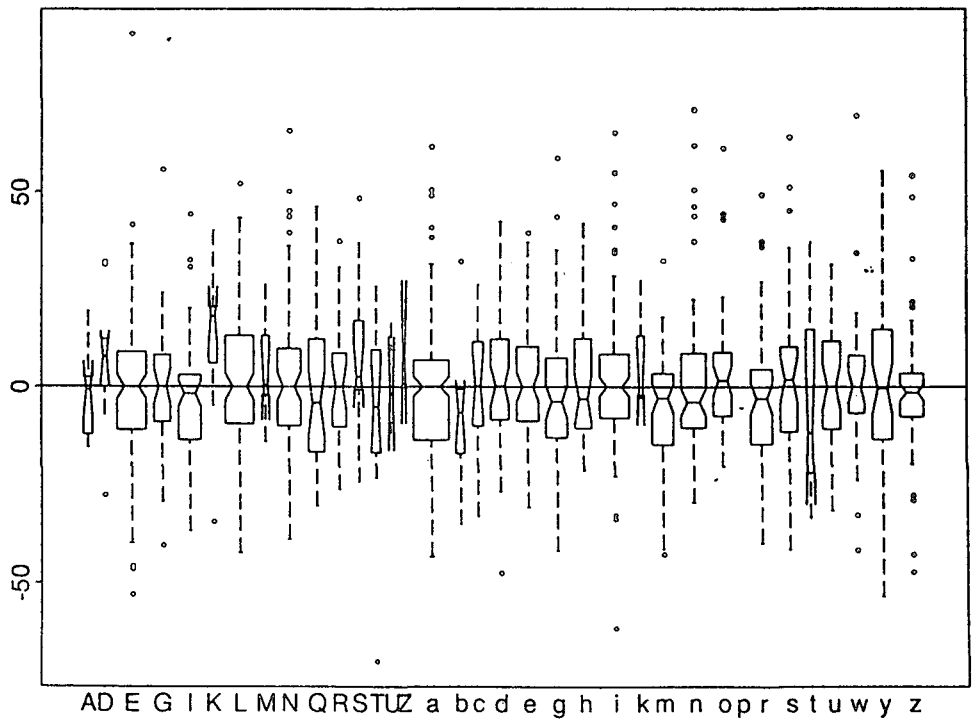

그림9. 각 음소의 지속시간 예측 에러