말소리와 음성과학 제3권 제2호 (2011)
pp. 23~28.

영어 발성에서 초음파 영상 정보를 이용한 인공신경망 기반의 인강부의 추정과 평가 방법에 대한 연구

Artificial Neural Network Prediction of Midsagittal Pharynx Shape from Ultrasound Images for English Speech

남호성[1)]

Hosung Nam

ABSTRACT

Electromagnetometers (EMA) have been widely used in articulatory studies as their temporal resolution can capture most speech activities and the fleshpoint information allows one to readily quantify and analyze tongue shape. However, the drawback is that the data lacks details of activity in the pharyngeal region. Several studies have attempted to estimate the unknown pharyngeal shape of the tongue, but few studies are based on unimodal data containing both front and back regions of the tongue at the same time. We use Stone's ball bearing method to obtain fleshpoint data as well as tongue shape. We further introduce a novel way of connecting balls and attaching them onto the tongue to ensure accurate tracking. An Artificial Neural Network is applied to build a map between observable flesh-points, unknown tongue shape, and pharyngeal region and is optimized to efficiently address nonlinearity.

Keywords: ANN, ultrasound imaging, pharynx estimation

1. 서론

발화(speech production) 할 때 조음(articulation)에 대한 연구는 그 기록하는 장치의 발전과 함께 발전해 왔다. X-ray, Electromagnetometer, MRI, 초음파 영상(ultrasound image)은 특히 구강에서 일어나는 발화, 특히, 가장 중요한 조음자(articulator)로서의 혀의 움직임에 대한 충실한 정보를 제공하기 때문에 다양한 조음관련 연구에서 널리 이용되어 왔다. 각각의 장치는 장단점을 동시에 가지고 있다. X-ray는 실험이 비교적 간편하고, 우수한 시공간 해상도를 제공하기 때문에 초기 연구에서 널리 이용되었지만, 방사선에 대한 노출을 감수해야 하는 위험 때문에, 최근에는 그 이용이 빈번하지 않다. 반면 정교한 영상을 제공하는 MRI는 보고된 위험이 없지만, 장비 이용료가 비싸며, 시간 해상도(temporal resolution)가 낮기 때문에 동적인 조음 연구에는 적합하지 않다. 최근에 실시간 MRI가 개발되었지만(Byrd et al 2009), 공간 해상도(spatial resolution)를 포기해야 하는 단점이 있다. 최근 들어 초음파 영상이 이 두 장비의 단점을 보완하며 각광받고 있다(Gick 2002 참조). 초음파 영상은 인체에 무해하고, 실험이 용이하며, 시공간적 해상도도 비교적 우수하다. 하지만, X-ray, MRI, 초음파 영상은 공통된 단점을 가지고 있다. 세 방법 모두 영상이 그 결과물이기 때문에 수량화된 분석이 쉽지 않다. 반면, Electromagnetometer(EMA 혹은 EMMA)는 뛰어난 시간 해상도와 함께, 영상이 아닌 좌표점(fleshpoint)이 그 결과값이므로 발화시 혀가 어떻게 움직이는지의 수량적 분석이 용이하다. 하지만, 다른 세 방법과는 달리 구강 전체의 형태 특히, 혀 뒤쪽(tongue back), 즉 인강부(pharynx)의 움직임을 기록할 수는 없다. 연구자들은 EMA와 같은 장치로 부터 얻어진 좌표점(fleshpoint) 정보로부터, 어떻게 기록되지 않은 인강부의 형태를 추정해 낼 수 있을까에 많은 관심을 가져 왔다(Kaburagi & Honda, 1994; Whalen et al., 1999; Magen et al., 2003; Jackson & McGowan, 2008). 하지만, 하나의 장비에서 좌표점 정보와 영상정보를 동시에 얻을 수 없기 때문에, 대부분의 연구는 서로 다른 장치의 결과물을 이용할 수 밖

1) Haskins Laboratories, nam@haskins.yale.edu

접수일자: 2011년 5월 13일
수정일자: 2011년 6월 8일
게재결정: 2011년 6월 11일

에 없었다(예: EMA로 부터의 좌표점 정보와 MRI와 같은 영상 정보를 비교). 그러한 다른 종류의 데이터 간의 정보추정에서 야기되는 오류를 최소화하기 위해, 본 연구는 영어 자음+모음 발화를 이용하여, 초음파 영상정보에서 좌표점값을 동시에 추출하는 방법을 소개하고자 한다. 또한, 혀표면의 한정된 수의 좌표점 값으로 부터, 기록되지 않은 인강부를 추정할 수 있는 인공신경망 기반 추정법을 제시한다. 마지막으로, 영어 CV의 동시조음(coarticulation) 효과를 추정된 결과와 함께 논의한다.

2. 데이터 수집

2.1 초음파 영상으로부터의 데이터 수집

Ziskin, et al.(1982)는 완벽한 구형의 단단한 물체가 그와 다른 음향적 저항(acoustic impedance)를 가진 물체(즉, 부드러운 물체) 위에 있을 때, 초음파 영상은 한 줄의 잔상(artifact tail)을 만들어 낸다는 것을 발견했다. Shawker et al.(1985)은 이러한 잔상을 3mm에서 5mm 직경의 쇠구슬 혹은 쇠공(ball bearing)을 혀에 놓았을 때 관찰할 수 있었고, Stone(1991)은 구체적인 실험방법과 함께 발화시 혀의 형태를 기록하는 좌표점 값을 구하는데 이용할 수 있음을 보여 줬다. Stone은 Shawker et al.이 이용한 쇠공을 이용하였고, 혀에 부착 시키기위해 치과 보철용 접착체 Impregum™을 이용했다. Impregum은 무해하고 초음파의 반사를 일으키지 않는다.

본 연구에서도 쇠공을 이용하여, 혀 표면의 중시상면(midsagittal) 형태와 좌표점(fleshpoint) 정보를 동시에 기록하고자 한다. 2.17, 3.17, 4, and 5.5mm 의 직경을 가진 쇠공이 테스트에 이용되었으며(각각은 0.8mm의 원통형 구멍이 그 중심을 통과한다), 4mm 직경의 쇠공이 발화의 용이함과 안정성을 고려하여 선택되었다.

Stone(1993)은 발화시 초음파 영상에서 좌표점(fleshpoint) 정보를 얻고자 할때, 어떻게 쇠공을 혀에 부착시켜야 하는지에 대한 방법을 제시한다. 하지만, 보고된 성공과는 달리, 우리는 몇 번의 pilot 실험에서 제시된 방법의 문제점을 발견하였다. 첫째, 사용된 치과보철용 접착제인 Impregum의 접착력 문제이다. 우리는 쇠공이 혀의 표면에 잘 부착되도록 하기 위해, 혀의 표면을 가능한 한 건조시켰다. 하지만, 쇠공을 상당 시간동안 안정적으로 혀표면에 유지하고 있기에는, 그 접착력이 충분하지 않았다. 실험 초반에 접착되어 있던 쇠공은 실험 도중 발화시 쉽게 떨어져 나갔다. 두 번째, 혀의 표면에 쇠공을 부착할 때, 원하는 곳에 쇠공을 위치시키기가 상당히 힘들다. 왜냐하면, 안전을 위해 여러 개의 공은 낚시 줄로 연결되어 있기 때문에, 하나의 공을 붙인 다음, 다른 하나의 공을 붙일때 그 공이 혀 위에 놓여 있게 된다. 이 때, Impregum의 특성상, 혀 위에 있는 공의 아래 부분에 그 접착제를 적절히 바른다는 것은 현실적으로 거의 불가능하다. 세 번째, Stone(1993)에서 소개된 Impregum은 그래서 우리가 기대한 만큼 초음파를 투과(sonolucent) 시키지는 못했다. 사실상 우리가 기대했던 잔상은(artifact tail)은 공과 혀 표면 사이에 아무것도 없을 때, 즉 Impregum이 없을 때만 관찰되었다. 그래서, 우리는 Impregum이 혀와 공을 서로 접착은 하지만, 혀와 공 사이에 Impregum이 최소한 도로 남아있게 할 필요가 있었다. 그러기 위해, Impregum을 공에 입힌 후, 혀에 접착할 때, 작은 나무 막대를 이용해 약간은 압력을 가해 주었다. 하지만, 이 땐 나무막대가 공에 붙어 버렸다.

이 모든 문제를 극복하기 위해선 새로운 형태의 접착제가 필요했고, 치과 교정용으로 쓰이고 있는 IsodentTM가 위에 언급된 대부분의 문제점을 효과적으로 해결해 줄 수 있다는 걸 알게 되었다. Isodent는 articulometer 연구에 널리 사용되고 있다. Isodent는 그 자체의 특성 때문에 접착강도, 지속성, 용이성이 모든 면에서 Impregum보다 뛰어났고, 걱정되었던 초음파 투과성은, 손에 직접 Isodent를 테스트해 봄으로서, Impregum보다 훨씬 우월함을 알 수 있었다.

위에 언급되었듯이 안전을 위해 낚시줄이 쇠공들을 관통하게 했다. 이 때, 공이 앞뒤로 자유로이 움직이는 걸 막기 위해 공의 양옆으로 매듭을 만들어 쇠공들 간의 거리가 일정하도록 유지했다. 하지만, 이 매듭은 공이 혀에 부착될 때, 낚시줄이 일직선으로 유지되는 것을 방해한다. 또한, 매듭은 쇠공사이의 거리를 고정시키는 효과가 있기 때문에, 발화시 혀가 움직이는 것을 부자연스럽게 할 뿐만 아니라, 혀가 움직임으로 인해서 야기되는 공사이 거리의 변화를 적절히 반영하지 못한다. 대안으로, 우리는 전기선을 절연시킬 때 사용하는 얇은 절연관을 3mm 씩 잘라, 공의 양쪽에 끼워 넣었다. 절연관은 열을 가하면, 수축되어 직경이 줄어든다. 적절한 마찰력이 줄에 생겨 공이 마음대로 움직이지는 못하지만, 힘을 가하면 어느 정도는 움직일 수 있을 정도로 수축시켰다. 낚시줄이 세 개의 쇠공을 관통하게 했고, 공 양쪽에는 적절히 수축된 절연관을 배치시켰다. 맨 앞의 쇠공은 혀 끝에서 2cm 뒤에, 나머지 두 공은 각각 1.5cm 뒤에 부착했다. 맨 앞의 쇠공은 턱뼈의 간섭때문에 실험 내내 관찰되지 않았다.

실험은 Aloka SSD5550 초음파 진단기를 통해 이루어졌으며, 자연스런 발화를 위해 약간의 머리 움직임 속에서도 신뢰할 수 있는 분석이 가능한 HOCUS(Haskins Optically Corrected Ultrasound System)를 이용하였다. Aloka SSD 5550 모델에서는 초당 127 frames의 영상을 얻을 수 있어서, 전동음(trill)같은 고속의 운동도 기록할 수 있다. 사용된 탐침자(probe)는 늑간(intercostal)에서 심장을 영상화할때 쓰이는 것으로, 7.5 MHz의 주파수를 이용한다. 공간 해상도는 약 1mm 정도이며, 시간 해상도는 8 msec이다.

기록된 2차원의 초음파 영상은 머리와 탐침자가 움직일 때마다 서로 상이한 좌표계을 갖게 된다. 이를 후처리로 보정할 수 있게 하기 위해, 머리와 탐침자의 3차원 좌표계의 변화를 Optotrak NDI 3020을 이용하여 기록하였다. 피험자의 머리엔 6개의 적외선 센서가 부착된 헬멧을 쓰게 했고, 초음파 탐침기에

는 5개의 센서를 부착하였다. Optotrak은 이 센서들의 움직임뿐만 아니라, 동시에 오디오 또한 디지털로 기록하였다. 초음파 영상 기록이 시작될 때, 클릭 소리가 Optotrak의 오디오 녹음시 추가 되었는데, 이것은 후에 초음파 영상과 Optotrak 데이타를 동기화 시키기 위함이다. 음성 녹음을 위해 피험자에게는 핸드프리 마이크(Audio-Technica Model ATM 75)가 장착되었다. 탐침자는 무게가 있어 안정되게 고정된 스탠드 위에 장착되었는데, 스프링이 탐침자에 부착되어 화자에 의한 상하 움직임에도 제자리를 찾을 수 있게 했다. 피시험자는 부채꼴의 형태의 탐침자가 턱밑에 세로로 위치하도록 앉았다. 피험자는 미국인 남성 영어화자이고, 18 자음+모음 (CV: [ə, s, k, r, l, st] + [i, a, u])의 형태를 발화했다 (ə는 자음이 아님에도 편의상 그렇게 했다). 각각의 발화는 3번씩 반복했다 (예외적으로 [əi, əu, əa]은 9번 [si, su, sa]은 6번 반복하였다).

2.2 데이터 처리

DICOM(Digital Imaging and Communication in Medicine) 포맷의 초음파 영상은 Matlab의 Image Processing Toolbox™ 을 이용해서 추출되었다. 수집된 음성 데이터를 참고하여, 각각의 CV 발화에서 C와 V의 중앙에 해당되는 시간을 찾고, 그 시점에서 혀 표면과 쇠공의 위치 정보를 EdgeTrak(Li et al. 2005)을 이용해 추출하여 171개의 데이터셋을 만들었다. 혀 표면의 윤곽 정보는 여러 개의 2차원 점들로 이루어져 있으며, 쇠공은 2개의 점 정보이다(세 개의 쇠공 중에 맨 앞의 쇠공은 턱뼈의 간섭으로 기록되지 않았다). 먼저, 혀 표면의 윤곽 정보는 수작업으로 임의의 수의 점정보로 기록된 후, 100 개의 점정보로 재배열(resampling) 된다. <그림 1> 는 [əa] 발화 중 [a]의 중간 시점에서 추출된 초음파 영상과 혀 표면 윤곽과 쇠공의 2차원 점이다.

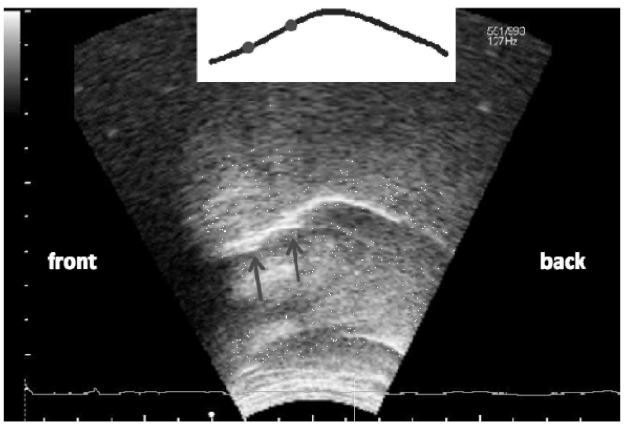

그림 1. [əa] 중 [a]의 초음파 영상 (아래)과 추출된 혀의 표면 윤곽과 쇠공의 점 정보 (위).

Fig. 1. An ultrasound image of [a] during [əa] (up); tongue surface contour and flesh point of tongue balls (down)

초음파 영상은 탐침자 기반 좌표계(probe-based coordinate)에서 수집되었기 때문에, 혀 윤곽과 쇠공의 점 정보 또한 탐침자 좌표계를 기반으로 한다. 이 탐침자 기반 정보는 일반적으로 조음 운동 연구에 이용되는 머리 좌표계(head-based coordinate)로 변환되어야 한다. 머리 좌표계는 화자의 머리를 참조하는 좌표계이므로, 조음 데이터가 다른 좌표계에서 수집되었을 땐, 음성학적으로 신뢰있는 분석을 하기 위해선 반드시 이 좌표계로 변환한 후 분석해야 한다. 머리 좌표계로의 변환은 Optotrak에서 얻어진, 헬멧과 탐침자 위의 적외선 센서의 3차원 위치 정보를 이용하여 이루어졌다.

본 연구는 인공신경망(Artificial Neural Network: ANN, 상세 내용은 3.1절 참조)을 이용하여, 혀에 부착된 한정된 수의 쇠공 정보로 혀 뒤편을 효과적으로 예측하고자 한다. 먼저, 171개의 데이타셋에 대해, 쇠공과 뒤혀 사이의 ANN 사상(mapping)에 필요한 입력과 출력을 정의할 필요가 있다. 입력은 두 쇠공의 x, y 점 정보이다. 앞서 언급했듯이, 혀끝에 위치한 공은 관찰되지 않았다. 여기서, 세 번째에 위치한 마지막 쇠공에서 혀 표면을 따라 1.5 cm 뒤 쪽에다 가상의 쇠공 위치정보를 추가로 입력 정보로 이용했다. 이 때, 가상의 쇠공이 세 번째 쇠공보다 1.5 cm 보다 더 멀어 질수도 가까와질 수 있다는 사실을 무시하였다. 이 가상의 쇠공의 추가함으로써, 쇠공으로 부터 뒤혀를 예측할 때 향상된 정확도를 기대할 수 있게 된다. 그래서 입력 정보로 2 혹은 3개의 쇠공의 위치 정보를 이용했다. 출력 정보는 쇠공 뒤의 혀의 표면의 윤곽 정보이다. 즉 많은 수의 점 정보이다. 이 때, 시작점과 끝점을 결정하는 것이 중요하다. 시작점은 혀 표면 위의 마지막 쇠구슬의 위치를 이용하였다. Surfaces (Parthasarathy et al., 2005)를 이용하여 혀 윤곽선의 꼬리부를 적당히 잘라 정돈하였고, 그 끝점이 출력점의 끝점이 되었다. 이 때, 출력부의 시작점에서 끝점까지의 점들은 등간격의 25개의 점들로 재배열(resampling) 하였다. 그래서, 입력과 출력의 차원은 4 x 50 혹은 6 x 50(한 점은 x, y로 이루어져 있기 때문에)이 된다. 한편 머리 좌표계로 변환된 정보는 그렇지 않은 원시 정보와 비교되었다. 입력 정보는 가상의 쇠공 설정여부에 따라, 출력 정보는 머리 좌표계로 보정 변환 되었는지 않았는지에 따라 구분했다.

3. 분석

3.1 인공신경망 (Artificial Neural Networks: ANN)

인공신경망(Artificial Neural Network: ANN)이 입출력(쇠공의 점정보로 부터 뒤혀의 윤곽선)의 매핑을 위해 선택되었다. 인공신경망은 많은 다른 인공 뉴런(artificial neuron)을 설정하여 분산적으로 적응시키는 학습 시스템이다. 각각의 뉴런은 외부로부터 입력을 받기도 하지만, 다른 뉴런으로 부터 신호를 전달받거나 자기 자신으로 부터 피드백을 받기도 한다. 뉴런들 간의

상호 연결된 방식은 신경망의 형태(topology)를 정의한다. 뉴런들 사이의 연결망을 통해 이루어지는 신호의 흐름은 각 연결선마다 부여된 연결강도에 의해 조절된다. 각각의 뉴런은 모든 다른 기여도의 신호들을 합하여 출력을 계산한다. 이 출력은 다시 다른 뉴런에 의해 처리되고, 결국은 시스템 전체의 출력이 된다. feed-forward 신경망은 전형적인 ANN이며, 하나 이상의 hidden layer를 갖는 동적 back-propagation(Rumelhart et al. 1986)에 의해 훈련된다. 이런 신경망은 사용이 쉽고, 어떤 형태의 입출력 매핑도 추정이 가능하다는 점에서, 함수의 추정, 패턴 분류, 인식 어플리케이션 등에 두루 사용되고 있다(Principe et al. 2000). 또한 신경망은 한 번 훈련이 되고 나면, 메모리와 실행 속도에 있어서 다른 방법에 비해 낮은 컴퓨팅 자원을 필요로 한다(Richmond 2001).

하나의 ANN을 훈련시킬 때, I의 입력과 T의 대상 출력을 필요로 한다. 훈련 단계에서, 각 뉴런의 연결강도(weight)와 바이어스(bias)는 신경망의 오차가 최소화될 때까지 반복적으로 수정된다. 전형적인 수행평가 기준으로 MSE(mean square error) - 입출력 사이 차이의 제곱에 대한 평균값 - 을 이용한다. ANN의 장점 중 하나는 다수의 입력과 다수의 출력에 대한 복잡한 매핑을 설정할 수 있다는 것이다. 이러한 구조에서는, 다수의 출력 대상들이 같은 hidden layers를 공유하기 때문에 출력 대상들 사이에 내재된 상호상관관계(cross-correlation)을 효과적으로 포착할 수 있다(Mitra et al. 2010). 본 연구에서 출력 정보가 될 위치 정보들은 어느 정도의 상호상관관계를 보일 것이 분명하므로, 우리는 ANN의 그러한 장점을 기대한다.

우리는 여러 시도를 거쳐, 하나의 hidden layer가 입력의 갯수(4 혹은 6)와 동일한 뉴런의 수를 가지는 것이 가장 적절하다는 것을 알게 되었다. Levenberg-Marquardt back-propagation 알고리즘이 optimization 규칙으로 이용되었고, 모든 layers에서 선형함수(linear function)가 여기(excitation)에 이용되었다. 입력과 대상 출력의 정보는 정규화(normalization)를 거친 후 ANN에 이용되었다.

3.2 데이터와 ANN의 보편성(Generality)

사용될 데이터(171개의 쇠공의 위치와 뒤 혀의 유곽선 정보)가 충분히 크지 않을 수도 있기 때문에, 두 가지 보편성에 대한 검증이 수반되어야 한다. 첫째, 수집된 데이터가 충분이 보편적인지 검증되어야 한다. 만약, 데이터가 수가 늘어났을 때, 다른 결과가 나온다면, 데이터의 보편성이 확보되지 못했다고 할 수 있다. 둘째, 훈련된 ANN이 데이터 전체의 보편적인 특징을 포착하고 있는지 검증되어야 한다. 예를 들어 하나의 ANN에 대해, 훈련(train)에 이용된 데이터만 잘 설명하고 검사(test)에 쓰인 데이터는 설명하지 못한다면, ANN의 보편성이 확보되지 않았다고 할 수 있다. 이런 경우를 과잉학습(overtraining)이라 한다.

이 두 보편성을 검증하기 위해, 두 가지 다른 ANN 실험을 했다. 각각의 ANN은 3.1 절에서 언급된 feedforward를 이용해서 훈련 및 검사를 했다. 첫째, 171개의 데이터셋에서 무작위로 75%를 선택한 후 훈련한 뒤 나머지 25%에서 검사했고, 이 과정을 100번 반복해서 실행하였다. 둘째, 훈련과 검사 데이터셋을 구분하지 않고, 데이타셋 전체를 훈련과 검사에 이용하였다. 각각의 실험은 2개의 쇠구슬 vs. 3개의 쇠구슬(가상의 구슬포함)로, 머리좌표계 보정과 무보정으로 구분된 데이터셋에서 독립적으로 실행되었다. 각 ANN 실험의 수행 정도는 RMSE(root mean square error)에 의해 평가되었다.

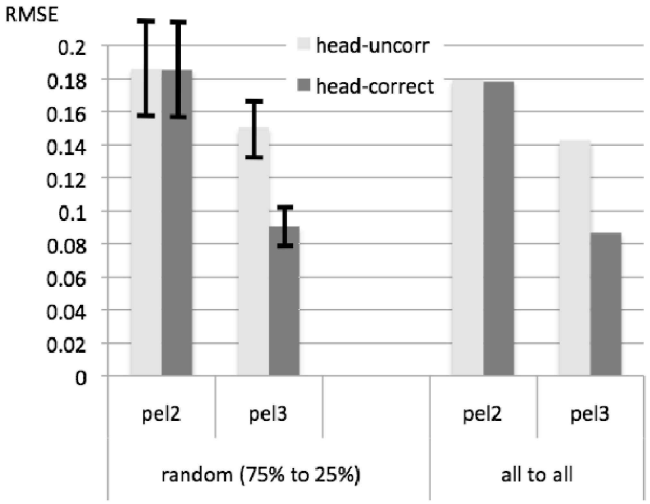

그림 2. 8개의 다른 실험의 RMSE 값: 무작위 75% 훈련(random 75% to 25%) 대 전체 훈련(all to all), 머리 보정(head-corr) 대 머리 무보정(head-uncorr), 2개(pel 2) 대. 3개의 쇠공(pel 3)
Fig. 2. RMSE values from 8 ANN experiments: random 75% vs. 100% for training; head-corrected vs. head-uncorrected; two pellets vs. three pellets.

<그림 2>는 8개의 다른 실험(무작위 75% 입력 대 전체 입력, 머리보정 대 머리 무보정, 2개 대 3개의 쇠공: 2 x 2 x 2)의 RMSE 값을 비교한다. 그래프에서 볼 수 있듯이, 무작위 75% 입력과, 전체 입력 실험의 양상은 전반적으로 아주 유사하다. 이것은 현재의 데이터셋과 ANN 훈련에 정의된 설정들이 혀의 기하학적 상관관계(혀 앞으로부터 혀 뒤로의 geometric correlation)의 보편성을 포착하기에 적절하다는 것을 말해준다. 나아가, 머리 보정의 효과는 가상의 쇠공이 추가 되었을 때만 나타났다. 이것은 두 개의 쇠공의 경우, 뒤혀의 추정 시 오차가 상대적으로 커서, 머리 보정의 효과를 상쇄시키기 때문이라고 볼 수 있다. 그리고, 이 결과는 머리 보정의 필요성을 동시에 설명하고 있다.

3.3 자음 모음의 동시조음(coarticulation effect)

이 절에선, 공정한 평가를 위해 전체 데이터를 이용하여 훈련(training)된 ANN을 이용하였다. 모음은 [i a u]로 분류하고,

자음은 뒤따르는 모음별 [i a u]로 구분하여 검사(testing)하고 그 RMSE를 비교하였다 <그림 3>. 앞선 보편성 실험에서처럼, 각각의 실험은 머리 보정의 유무, 가상의 쇠공 사용 유무에 대해 독립해서 이루어졌다. 그래프에서 볼 수 있듯이, 모음에서 보이는 RMSE 패턴은 그 모음을 선행하는 자음에서도 발견되었다. 즉, 가장 높은 RMSE는 [i]에서 그리고, [i]앞의 자음에서 찾아 볼 수 있다. 특히, 이러한 RMSE 패턴은 머리 보정과 가상 쇠공의 추가 이후 더 명확히 드러난다. 아마도 이러한 RMSE 패턴은 모음의 조음이 선행하는 자음에서도 일어나는 동시조음 효과로 볼 수 있다. Löfqvist & Gracco(1994)는 영어 조음 연구에서, 음절말(syllable final 혹 coda)과는 달리 음절초(syllable initial 혹은 onset)에서 자음과 모음이 동시에 조음됨을 보였다. 음절말에서도 이런 RMSE 패턴이 발견되는지에 대한 연구가 뒤따를 필요가 있다.

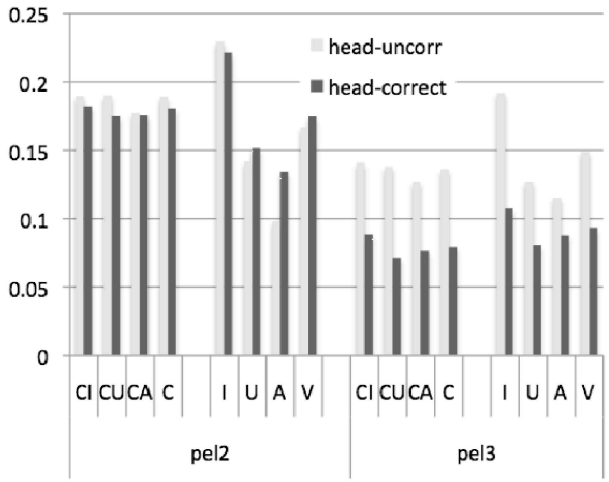

그림 3. 모음과 후행모음의 유형별로 구분된 자음의 ANN 수행의 RMSE 값: CI-[i]앞의 자음, CU-[u]앞의 자음, CA-[a]앞의 자음. C 와 V는 자음 모음 전체에 대한 평균 RMSE.

Fig. 3. RMSE values from ANN experiments for different vowels and consonants with different vowel contexts: CI – consonants before [i]; CU – consonants before [u]; CA – consonants before [A]; C and V are average RMSE for all consonants and vowels, respectively.

4. 결론

인공 신경망은(ANN) 혀 앞의 제한된 정보를 이용하여 혀 뒤를 적절히 예측하는 추정 툴로서의, 그리고, 머리 보정과 자모음 동시 조음을 위한 음성/음운적 분석 툴로서의 가능성을 제시했다. 또한, 초음파를 발화 연구에 이용할 때, 머리 보정의 중요성을 보여 주었다. 턱뼈의 간섭으로 사라진 쇠공의 위치 정보는 가상의 쇠공을 추가함으로서 어느 정도 ANN의 수행을 향상시킬 수 있었다.

감사의 글

이 연구는 NIH DC-02717의 지원 하에 이루어졌고, Vikramjit Mitra, Khalil Iskarus, Doug Whalen, Anthony De Simone, Aude Noiray, Maureen Stone, Mark Tiede, 마지막으로 익명의 심사위원들에게 감사를 표한다.

참고문헌

Byrd, D., Tobin, S., Bresch, E., & Narayanan, S. (2009). “Timing effects of syllable structure and stress on nasals: a real-time MRI examination”, *Journal of Phonetics*, Vol. 37, 97-110.

Gick, B. (2002). “The use of ultrasound for linguistic phonetic fieldwork”, *Journal of the International Phonetic Association*, Vol. 32, No. 2, 113-121.

Jackson, M. T. T., & McGowan, R. S. (2008). “Predicting midsagittal pharyngeal dimensions from measures of anterior tongue position in Swedish vowels: Statistical considerations”, *Journal of the Acoustical Society of America*, Vol. 123, No. 1, 336-346.

Kaburagi, T., & Honda, M. (1994). “An ultrasonic method for monitoring tongue shape and the position of a fixed-point on the tongue surface”, *Journal of the Acoustical Society of America*, Vol. 95, No. 4, 2268-2270.

Kaburagi, T., & Honda, M. (1994). “Determination of sagittal tongue shape from the positions of points on the tongue Surface”, *Journal of the Acoustical Society of America*, Vol. 96, No. 3, 1356-1366.

Li, M., Kambhamettu, C., & Stone, M. (2005). “Automatic contour tracking in ultrasound images”, *Clinical Linguistics and Phonetics*, Vol. 19, No. 6-7, 545-554.

Löfqvist, A., & Gracco, V. L. (1994). “Tongue body kinematics in velar stop production: influences of consonant voicing and vowel context”, *Phonetica*, Vol. 51, No. 1-3, 52-67.

Magen, H. S., Kang, A. M., Tiede, M. K., & Whalen, D. H. (2003). “Posterior pharyngeal wall position in the production of speech”, *Journal of Speech, Language, and Hearing Research*, Vol. 46, 241-251.

Mitra, V., Nam, H., Espy-Wilson, C., Saltzman, E., & Goldstein, L. (2010). “Retrieving tract variables from acoustics: a comparison of different Machine Learning strategies”, *The IEEE Journal of Selected Topics in Signal Processing*, Vol. 4, No. 6, 1027-1045.

Parthasarathy, V., Stone M., & Prince, J.L. (2005). “Spatiotemporal visualization of the tongue surface using ultrasound and kriging (SURFACES)”, *Clinical Linguistics and Phonetics*, Vol. 19, No.

6-7, 529-544.

Principe, C., Euliano, N. R., & Lefebvre, W. C. (2000). *Neural and Adaptive Systems: Fundamentals Through Simulations*. New York: Wiley.

Richmond, K. (2001). "Estimating articulatory parameters from the Speech Signal", Ph.D. Thesis, University of Edinburgh.

Rumelhart, D.E., Hinton, G.E. & Williams, R.J. (1986). "Learning internal representations by error propagation", *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, 318-362, Cambridge, MA: MIT Press.

Shawker, T. H., & Sonies, B. C. (1984). "Tongue movement during Speech - a real-time ultrasound evaluation", *Journal of Clinical Ultrasound*, Vol. 12, No. 3, 125-133.

Stone, M. (1991). "Toward a three dimensional model of tongue movement", *Journal of Phonetics*, Vol. 19, 309-320.

Stone, M. (1993). ASA speech production video.

Whalen, D. H., Kang, A. M., Magen, H. S., Fulbright, R. K., & Gore, J. C. (1999). "Predicting midsagittal pharynx shape from tongue position during vowel production", *Journal of Speech Language and Hearing Research*, Vol. 42, No. 3, 592-603.

Ziskin, C., Thickman, V. & Goldenberg J. (1982). "The comet tail artifact", *Journal of Ultrasound in Medicine*, Vol. 1, 1-7.

• **남호성 (Nam, Hosung)**
Haskins Laboratories
300 George st. New Haven, CT 06511, USA
Tel: 1-203-865-6163 (ext. 235)
Email: nam@haskins.yale.edu
관심분야: 음성학, 음운론, 음성공학, 언어습득
현재 Research Scientist로 재직 중