



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0095603
(43) 공개일자 2017년08월23일

(51) 국제특허분류(Int. Cl.)
G10L 15/22 (2006.01) A61B 5/0488 (2006.01)
G06K 9/00 (2006.01) G10L 15/02 (2006.01)
G10L 15/18 (2006.01)

(52) CPC특허분류
G10L 15/22 (2013.01)
A61B 5/0488 (2013.01)

(21) 출원번호 10-2016-0017213

(22) 출원일자 2016년02월15일

심사청구일자 2016년02월15일

(71) 출원인
인하대학교 산학협력단

인천광역시 남구 인하로 100, 인하대학교 (용현동)

(72) 발명자
김덕환

서울특별시 양천구 목동로 230, 103동 1004호 (목동, 아이파크아파트)

이병현

인천광역시 남동구 구월로 45-9, 1층 (간석동)

(74) 대리인

특허법인 정안

전체 청구항 수 : 총 10 항

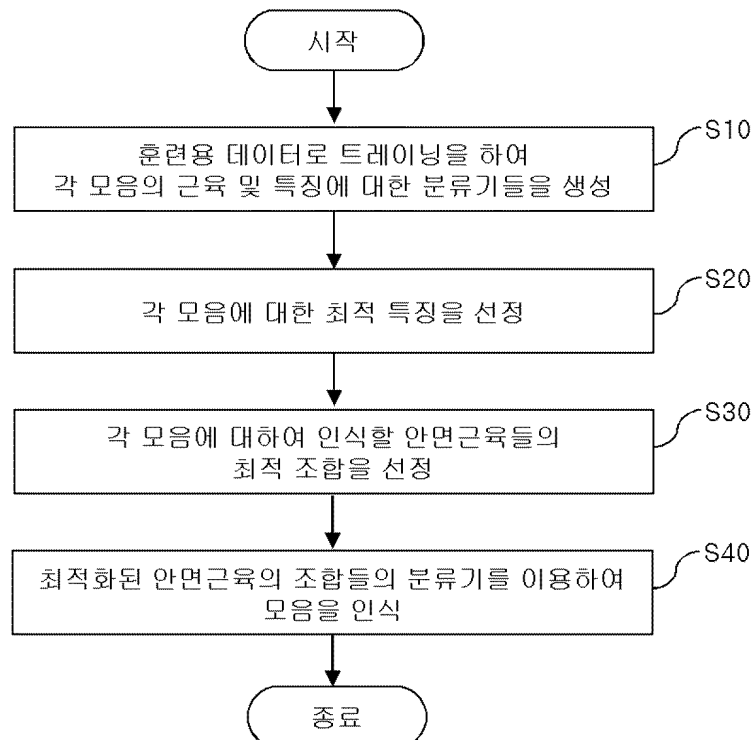
(54) 발명의 명칭 근육 조합 최적화를 통한 안면근육 표면근전도 신호기반 단모음인식 방법

(57) 요약

본 발명은 안면근육의 표면근전도 신호를 기반으로 분류기를 적용하여, 발생 없이 근전도 신호만으로 한국어 단모음을 인식하는, 안면근육 표면근전도 신호기반 단모음인식 방법에 관한 것으로서, (a) 모음을 발음할 때 측정되는 다수의 안면근육의 표면근전도 신호의 훈련용 데이터를 이용하여, 각 모음의 안면근육 및 특징에 의한 분류

(뒷면에 계속)

대표도 - 도2



기들을 생성하는 단계; (b) 각 모음의 각 안면근육에 대하여 사용할 특징을 선정하는 단계; (c) 각 모음을 인식하는 안면근육들의 분류기 조합들 중 해당 모음을 인식할 분류기 조합을 선정하는 단계; 및, (d) 선정된 분류기 조합으로 해당 모음을 인식하는 단계를 포함하는 구성을 마련한다.

상기와 같은 방법에 의하여, 발성 없이도 안면근육의 근전도만으로 모음을 인식함으로써, 원하지 않는 음성 노이즈에 노출되지 않기 때문에 옥외 장소나 차내 등 소음이 많은 곳에서도 정확하게 음성을 인식할 수 있고, 발성자가 성대에 손상이 있더라도 음성인식을 할 수 있다.

(52) CPC특허분류

G06K 9/00885 (2013.01)

G10L 15/02 (2013.01)

G10L 15/18 (2013.01)

명세서

청구범위

청구항 1

안면근육 표면근전도 신호기반 단모음인식 방법에 있어서,

- (a) 모음을 발음할 때 측정되는 다수의 안면근육의 표면근전도 신호의 훈련용 데이터를 이용하여, 각 모음의 안면근육 및 특징에 의한 분류기들을 생성하는 단계;
- (b) 각 모음의 각 안면근육에 대하여 사용할 특징을 선정하는 단계;
- (c) 각 모음을 인식하는 안면근육들의 분류기 조합들 중 해당 모음을 인식할 분류기 조합을 선정하는 단계; 및,
- (d) 선정된 분류기 조합으로 해당 모음을 인식하는 단계를 포함하는 것을 특징으로 하는 안면근육 표면근전도 신호기반 단모음인식 방법.

청구항 2

제1항에 있어서,

상기 (b)단계에서, 각 모음 마다 각 안면근육 별로 다수의 특징들 중 가장 높은 인식 정확도를 가지는 특징으로 선정하는 것을 특징으로 하는 안면근육 표면근전도 신호기반 단모음인식 방법.

청구항 3

제1항에 있어서,

상기 (c)단계에서, 각 모음 별로 인식 정확도가 가장 높은 순으로 안면근육의 분류기들을 조합하여, 인식 정확도가 가장 높은 조합을 해당 분류기 조합으로 선정하는 것을 특징으로 하는 안면근육 표면근전도 신호기반 단모음인식 방법.

청구항 4

제1항에 있어서,

상기 (c)단계에서, 다수의 분류기 조합은 조합된 분류기의 결과들을 논리식으로 결합하여 모음의 인식 결과를 분류하는 것을 특징으로 하는 안면근육 표면근전도 신호기반 단모음인식 방법.

청구항 5

제4항에 있어서,

상기 (c)단계에서, 다수의 분류기 조합은 조합된 분류기의 적어도 50% 이상의 분류기가 해당 모음을 인식하는 경우 해당 모음을 인식하는 것으로 분류하는 것을 특징으로 하는 안면근육 표면근전도 신호기반 단모음인식 방법.

청구항 6

제1항에 있어서,

상기 표면근전도 신호는 악이복근 전복(Anterior Belly of Digastricus), 입꼬리올림근(Levator Anguli Oris), 대관골근(Zygomaticus Major), 광경근(Platysma), 교근(Masseter) 중 적어도 1개 이상의 안면근육으로부터 검출되는 신호인 것을 특징으로 하는 안면근육 표면근전도 신호기반 단모음인식 방법.

청구항 7

제1항에 있어서,

상기 모음은 한국어 단모음으로서, 'ㅏ', 'ㅑ', 'ㅓ', 'ㅕ', 'ㅗ', 'ㅛ', 'ㅜ', 'ㅠ' 중 적어도 1개 이상인 것을 특징으로 하는 안면근육 표면근전도 신호기반 단모음인식 방법.

청구항 8

제1항에 있어서,

상기 특징은 RMS(Root Mean Square), VAR(Variance), MMAV1(Modified Mean Absolute Value 1), MMAV2(Modified Mean Absolute Value 2), 캡스트럼 계수(Cepstral Coefficients) 중 어느 하나 이상의 특징 추출 알고리즘에 의해 구해지는 것을 특징으로 하는 안면근육 표면근전도 신호기반 단모음인식 방법.

청구항 9

제1항에 있어서,

상기 분류기는 QDA(Quadratic Discriminant Analysis) 또는 HMM(Hidden Markov Model) 분류기인 것을 특징으로 하는 안면근육 표면근전도 신호기반 단모음인식 방법.

청구항 10

제1항 내지 제9항 중 어느 한 항의 안면근육 표면근전도 신호기반 단모음인식 방법을 수행하는 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

발명의 설명

기술 분야

[0001] 본 발명은 안면근육의 표면근전도 신호를 기반으로 분류기를 적용하여, 발성 없이 근전도 신호만으로 한국어 단모음을 인식하는, 안면근육 표면근전도 신호기반 단모음인식 방법에 관한 것이다.

배경 기술

[0002] 일반적으로, 자동음성인식(Automatic Speech Recognition)은 컴퓨터가 음성 신호를 받아서 발화자의 말을 인식하게 하는 기술이다. 자동음성인식에 대한 연구가 심화됨에 따라 성공적으로 전화 기반 서비스나 모바일 서비스 등에 적용되었다[비특허문헌 1,2]

[0003] 자동음성인식은 높은 인식 정확도를 보이지만 여러 가지 애로사항이 있다. 먼저, 자동음성인식은 공기를 매개체로 하여 전달되는 음성 신호의 입력을 전제로 한다. 이 때문에 원하지 않는 음성 노이즈에 노출되어 있으며, 노이즈에 의해 상당한 영향을 받아 인식 정확도가 떨어진다. 이것은 옥외, 공공장소, 또는 차내가 될 수 있다[비특허문헌 3,4]. 둘째로, 공공장소에서 음성인식 사용으로 인한 개인정보 노출의 위험이 있으며, 셋째로 사무실이나 도서관에서 원하지 않는 소음이 될 수 있다[비특허문헌 5,6]. 넷째로 기관절개술, 후두적출 등 성대에 손상을 주는 수술을 받은 사람과 구음장애, 실어증 등 발성에 영향을 주는 장애를 가진 사람은 자동음성인식을 사용하는데 한계가 있다[비특허문헌 7]. 마지막으로 사고나 자연재해, 테러나 범죄 같은 위급한 상황에서 소리를

낼 수 없거나 음성 노이즈가 심한 상황에서의 자동음성인식 사용은 제한된다[비특허문헌 8].

[0004]

이러한 애로사항을 보완 및 대체하기 위한 연구들이 진행되고 있다[비특허문헌 3,5-8]. 이 연구들은 조음(Articulation)에 사용되는 안면근육의 근전도 신호를 이용하여 음성 인식을 하는 것으로, 음성 노이즈에 영향을 거의 받지 않는다[비특허문헌 3,5-8]. 또한, 자동음성인식은 발음이 비슷한 소리를 구분하는데 어려움이 있지만, 근전도 신호를 이용한 방법은 상대적으로 구분하기 쉽다. 이는 조음 방법이 다르면 사용하는 근육이 달라지기 때문이다[비특허문헌 9].

선행기술문헌

비특허문헌

[0005]

(비특허문헌 0001) [비특허문헌 1] Apple Siri, <http://www.apple.com/ios/siri/>

(비특허문헌 0002) [비특허문헌 2] Google Voice Actions, <https://developers.google.com/voice-actions/>

(비특허문헌 0003) [비특허문헌 3] S. Kumar, D. K. Kumar, M. Alemu, M. Berry, "EMG Based Voice Recognition", in Proc. of IEEE Conf. on Intelligent Sensors, Sensor Network and Information Processing, pp. 593-597, Melbourne, Australia, Dec 2004.

(비특허문헌 0004) [비특허문헌 4] J. F. Gemmeke, T. Virtanen, A. Hurmalainen, "Exemplar-Based Sparse Representations for Noise Robust Automatic Speech Recognition", IEEE Trans. Audio, Speech and Language Processing, Vol. 19, no. 7, Sep. 2011.

(비특허문헌 0005) [비특허문헌 5] T. Heistermann, M. Janke, M. Wand, T. Schultz, "Spatial Artifact Detection for Multi-Channel EMG-Based Speech Recognition", International Conf. on Bio-inspired Systems and Signal Processing, pp. 189-196, Angers, France, Mar 2014.

(비특허문헌 0006) [비특허문헌 6] H. Manabe, Z. Zhang, "Multi-stream HMM for EMG-based speech recognition", in Proc. of IEEE Conf. on Engineering in Medicine and Biology Society, pp. 4389-4392, San Francisco, CA, Jun 2004.

(비특허문헌 0007) [비특허문헌 7] Y. Deng, R. Patel, J. T. Heaton, G. Colby, L. D. Gilmore, J. Cabrera, S. H. Roy, C. J. D. Luca, G. S. Meltzner, "Disordered speech recognition using acoustic and sEMG signals", INTERSPEECH 2009, pp. 644-647, Brighton, UK, Sep 2009.

(비특허문헌 0008) [비특허문헌 8] C. Jorgensen, S. Dusan, "Speech interfaces based upon surface electromyography", Speech Communication, Vol. 20, no. 4, pp. 354-366, Apr 2010.

(비특허문헌 0009) [비특허문헌 9] A.D.C. Chan, K. Englehart, B. Hudgins, D.F. Lovely, "Hidden Markov Model Classification of Myoelectric Signals in Speech", IEEE Trans. Engineering in Medicine and Biology Magazine, Vol. 21, no. 4, pp. 143-146, Sep 2002.

(비특허문헌 0010) [비특허문헌 10] H. Yong, "A Typological Study on Korean Vowel Systems", Language and Linguistics, pp. 175-200, Vol. 61, Nov 2013.

(비특허문헌 0011) [비특허문헌 11] A. Phinyomark, S. Hirunviriya, C. Limsakul, P. Phukpattaranont, "Evaluation of EMG Feature Extraction for Hand Movement Recognition Based on Euclidean Distance and Standard Deviation", in Proc. of IEEE Conf. on ECTI, pp.856-860, Chiang Mai, Thailand, May 2010.

(비특허문헌 0012) [비특허문헌 12] E. Scheme, K. Englehart, "On the Robustness of EMG Features for Pattern Recognition Based Myoelectric Control; A Multi-Dataset Comparison", in Proc. of IEEE Conf. on EMBS, pp.650-653, Chicago, USA, Aug 2014.

(비특허문헌 0013) [비특허문헌 13] B-H. Lee, J-H. Ryu, M-R. Lee, S-H. Kim, M. Z. Uddin, D-H. Kim, "Monophthong recognition using feature and muscle selection based on facial surface EMG signals", in Proc. of The IEEE Conf. on Summer Conference, pp. 933-936, Jeju, Korea, Jun 2015.

(비특허문헌 0014) [비특허문헌 14] N. Srisuwan, P. phukpattaranont, C. Limsakul, "Three Steps of Neuron Network Classification for EMG-based Thai Tones Speech Recognition", in Proc. of IEEE Conf. on

ECTI, pp. 1-6, Krabi, Thailand, May 2013.

(비특허문헌 0015) [비특허문헌 15] E. Lopez-Larraz, O. M. Mozos, J. M. Antelis, J. Minguez, "Syllable-Based Speech Recognition Using EMG", in Proc. of IEEE Conf. on EMBS, pp. 4699-4702, Buenos Aires, Argentina, Aug 2010.

발명의 내용

해결하려는 과제

- [0006] 본 발명의 목적은 상술한 바와 같은 문제점을 해결하기 위한 것으로, 안면근육의 표면근전도 신호를 기반으로 분류기를 적용하여, 발성 없이 근전도 신호만으로 한국어 단모음을 인식하는, 안면근육 표면근전도 신호기반 단모음인식 방법을 제공하는 것이다.
- [0007] 특히, 본 발명의 목적은 각 피험자의 단모음 발음에 따라 활성도가 높은 근육과 인식률이 높은 특징을 추출하여 근육 조합을 최적화하고, 최적화된 근육 조합으로 한국어 단모음을 분류하여 인식하는, 안면근육 표면근전도 신호기반 단모음인식 방법을 제공하는 것이다.
- [0008] 또한, 본 발명의 목적은 근육 조합 최적화를 적용하고, 기계학습과 통계적 분류기로 사용되는 QDA(Quadratic Discriminant Analysis)와 기존의 음성인식 시스템에서 통계적 모델로 사용되는 HMM(Hidden Markov Model)을 분류기로 이용하여 한국어 단모음을 인식하는, 안면근육 표면근전도 신호기반 단모음인식 방법을 제공하는 것이다.
- [0009]

과제의 해결 수단

- [0010] 상기 목적을 달성하기 위해 본 발명은 안면근육 표면근전도 신호기반 단모음인식 방법에 관한 것으로서, (a) 모음을 발음할 때 측정되는 다수의 안면근육의 표면근전도 신호의 훈련용 데이터를 이용하여, 각 모음의 안면근육 및 특징에 의한 분류기들을 생성하는 단계; (b) 각 모음의 각 안면근육에 대하여 사용할 특징을 선정하는 단계; (c) 각 모음을 인식하는 안면근육들의 분류기 조합들 중 해당 모음을 인식할 분류기 조합을 선정하는 단계; 및, (d) 선정된 분류기 조합으로 해당 모음을 인식하는 단계를 포함하는 것을 특징으로 한다.
- [0011] 또, 본 발명은 안면근육 표면근전도 신호기반 단모음인식 방법에 있어서, 상기 (b)단계에서, 각 모음마다 각 안면근육 별로 다수의 특징들 중 가장 높은 인식 정확도를 가지는 특징으로 선정하는 것을 특징으로 한다.
- [0012] 또, 본 발명은 안면근육 표면근전도 신호기반 단모음인식 방법에 있어서, 상기 (c)단계에서, 각 모음 별로 인식 정확도가 가장 높은 순으로 안면근육의 분류기들을 조합하여, 인식 정확도가 가장 높은 조합을 해당 분류기 조합으로 선정하는 것을 특징으로 한다.
- [0013] 또, 본 발명은 안면근육 표면근전도 신호기반 단모음인식 방법에 있어서, 상기 (c)단계에서, 다수의 분류기 조합은 조합된 분류기의 결과들을 논리식으로 결합하여 모음의 인식 결과를 분류하는 것을 특징으로 한다.
- [0014] 또, 본 발명은 안면근육 표면근전도 신호기반 단모음인식 방법에 있어서, 상기 (c)단계에서, 다수의 분류기 조합은 조합된 분류기의 적어도 50% 이상의 분류기가 해당 모음을 인식하는 경우 해당 모음을 인식하는 것으로 분류하는 것을 특징으로 한다.
- [0015] 또, 본 발명은 안면근육 표면근전도 신호기반 단모음인식 방법에 있어서, 상기 표면근전도 신호는 악이복근 전복(Anterior Belly of Digastricus), 입꼬리올림근(Levator Anguli Oris), 대관골근(Zygomaticus Major), 광경근(Platysma), 교근(Masseter) 중 적어도 1개 이상의 안면근육으로부터 검출되는 신호인 것을 특징으로 한다.
- [0016] 또, 본 발명은 안면근육 표면근전도 신호기반 단모음인식 방법에 있어서, 상기 모음은 한국어 단모음으로서, 'ㅏ', 'ㅑ', 'ㅓ', 'ㅕ', 'ㅗ', 'ㅛ', 'ㅜ', 'ㅠ' 중 적어도 1개 이상인 것을 특징으로 한다.
- [0017] 또, 본 발명은 안면근육 표면근전도 신호기반 단모음인식 방법에 있어서, 상기 특징은 RMS(Root Mean Square), VAR(Variance), MMAV1(Modified Mean Absolute Value 1), MMAV2(Modified Mean Absolute Value 2), 캡스트럼 계수(Cepstral Coefficients) 중 어느 하나 이상의 특징 추출 알고리즘에 의해 구해지는 것을 특징으로 한다.

[0018] 또, 본 발명은 안면근육 표면근전도 신호기반 단모음인식 방법에 있어서, 상기 분류기는 QDA(Quadratic Discriminant Analysis) 또는 HMM(Hidden Markov Model) 분류기인 것을 특징으로 한다.

[0019] 또한, 본 발명은 안면근육 표면근전도 신호기반 단모음인식 방법을 수행하는 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것이다.

발명의 효과

[0020] 상술한 바와 같이, 본 발명에 따른 안면근육 표면근전도 신호기반 단모음인식 방법에 의하면, 발생 없이도 안면근육의 근전도만으로 모음을 인식함으로써, 원하지 않는 음성 노이즈에 노출되지 않기 때문에 옥외 장소나 차내 등 소음이 많은 곳에서도 정확하게 음성을 인식할 수 있고, 발생자가 성대에 손상이 있더라도 음성인식을 할 수 있는 효과가 얻어진다.

도면의 간단한 설명

[0021] 도 1은 본 발명의 일실시예에 따른 안면근육 표면근전도 신호기반 단모음인식 방법을 실시하기 위한 전체 시스템의 구성에 대한 블록도.

도 2는 본 발명의 일실시예에 따른 안면근육 표면근전도 신호기반 단모음인식 방법을 설명하는 흐름도.

도 3은 본 발명의 일실시예에 따른 안면근육 표면근전도 신호기반 단모음인식 방법을 설명하는 상세 흐름도.

도 4는 본 발명의 일실시예에 따른 데이터 수집에 사용된 안면근육을 나타낸 도면.

도 5는 본 발명의 일실시예에 따른 한국어 단모음에 따른 (a) 캡스트럼 계수(Cepstral Coefficients)와 (b) RMS 특징값을 나타낸 그래프.

도 6은 본 발명의 일실시예에 따른 QDA를 통한 'ㅏ', 'ㅑ', 'ㅣ' 발음의 캡스트럼 계수(Cepstral Coefficients) 특징값 분류를 나타낸 그래프.

도 7은 본 발명의 일실시예에 따른 음성인식에 사용되는 HMM 모델을 나타낸 도면.

도 8은 본 발명의 실험에 따른 실험에 사용된 장비를 촬영한 이미지.

도 9는 본 발명의 실험 결과에 따른 QDA를 사용한 피험자 3과 피험자 8의 근육에 따른 단모음별 인식 정확도를 나타낸 그래프.

도 10은 본 발명의 실험 결과에 따른 전체 피험자의 근육별 사용된 특징 횟수를 나타낸 그래프.

도 11은 본 발명의 실험 결과에 따른 전체 피험자의 근육 조합별 및 근육 최적화 인식 정확도를 나타낸 그래프로써, (a) QDA, (b) HMM를 나타낸 그래프.

도 12는 본 발명의 실험 결과에 따른 전체 피험자의 한국어 단모음 인식 혼동행렬(Confusion matrix)을 나타낸 표로써, (a) QDA, (b) HMM를 나타낸 표.

도 13은 본 발명의 실험 결과에 따른 다른 연구와의 평균 인식 정확도 비교를 나타낸 표.

발명을 실시하기 위한 구체적인 내용

[0022] 이하, 본 발명의 실시를 위한 구체적인 내용을 도면에 따라서 설명한다.

[0023] 또한, 본 발명을 설명하는데 있어서 동일 부분은 동일 부호를 붙이고, 그 반복 설명은 생략한다.

[0024] 먼저, 본 발명의 일실시예에 따른 안면근육 표면근전도 신호기반 단모음인식 방법을 실시하기 위한 전체 시스템의 구성에 대하여 도 1을 참조하여 설명한다.

[0025] 도 1에서 보는 바와 같이, 본 발명에 따른 안면근육 표면근전도 신호기반 단모음인식 방법은 훈련 데이터(11) 또는 근전도 신호(12)를 입력받아, 특징 및 채널을 선택하거나 사용자를 인식하는 컴퓨터 단말(20) 상의 프로그램 시스템(30)으로 실시될 수 있다. 즉, 상기 안면근육 표면근전도 신호기반 단모음인식 방법은 프로그램으로 구성되어 컴퓨터 단말(20)에 설치되어 실행될 수 있다. 컴퓨터 단말(20)에 설치된 프로그램은 하나의 프로그램 시스템(30)과 같이 동작할 수 있다.

[0026] 한편, 다른 실시예로서, 상기 안면근육 표면근전도 신호기반 단모음인식 방법은 프로그램으로 구성되어 범용 컴

퓨터에서 동작하는 것 외에 ASIC(주문형 반도체) 등 하나의 전자회로로 구성되어 실시될 수 있다. 또는 안면근육의 근전도 신호를 이용한 사용자의 음성 인식만을 전용으로 처리하는 전용 단말(30)로 개발될 수도 있다. 특히, 상기와 같은 전자회로는 사용자를 인식하는 인식장치에 이용되거나, 인식장치의 일부로서 실시될 수 있다. 이를 표면근전도 신호기반 단모음인식 장치(30)라 부르기로 한다. 그 외 가능한 다른 형태도 실시될 수 있다.

[0027] 다음으로, 본 발명에 따라 사용자의 음성 인식을 할 때 이용되는 단모음 체계, 및, 안면 근육에 대하여 설명하고, 종래기술에 따른 근전도 신호에 기반한 음성인식 방법에 대하여 설명한다.

[0028] 먼저, 한국어 단모음 체계에 대하여 설명한다.

[0029] 현대 한국어의 단모음은 'ㅏ', 'ㅑ', 'ㅓ', 'ㅕ', 'ㅗ', 'ㅛ', 'ㅜ', 'ㅠ', 'ㅡ', 'ㅣ' 모두 10개이다. 이를 전통적인 단모음체계, 또는 10모음체계라고 부른다. 하지만 'ㅑ'는 'ㅓ'와 구분되기 힘들며, 'ㅛ', 'ㅠ'를 이중모음으로 발음 할 경우를 고려하면 실제로 사용하는 단모음체계는 7모음체계가 된다[비특허문헌 10]. 따라서 본 발명에서는 실생활에서 사용되는 7모음체계를 기준으로 단모음을 인식하였다.

[0030] 다음으로, 안면근육에 대하여 설명한다.

[0031] 근전도 신호를 이용한 음성인식에 관한 이전 연구들 중에서 근전도 신호를 위한 전극 위치는 경험적으로 악이복근 전복(Anterior Belly of Digastricus), 입꼬리올림근(Levator Anguli Oris), 입꼬리내림근(Depressor Anguli Oris), 대관골근(Zygomaticus Major), 광경근(Platysma)[비특허문헌 5,9], 입둘레근(Orbicularis Oris)[비특허문헌 6], 교근(Masseter), 턱끝근(Mentalis)[비특허문헌 3] 이 결정되었다. 또한, 근전도 신호를 수집하기 위해서 침습형 전극을 사용하는 것은 사용자에게 불편함을 주기 때문에 비침습형 전극을 사용한 표면 근전도 신호 수집이 주로 이루어졌다.

[0032] 바람직하게는, 본 발명에서는, 악이복근 전복, 입꼬리올림근, 대관골근, 교근, 광경근 등 5개의 안면근육에서 발생하는 근전도 신호를 사용한다.

[0033] 다음으로, 종래기술에 따른 근전도 신호기반 음성인식 방법에 대하여 설명한다.

[0034] 근전도 신호를 이용한 음성인식은 발성과 함께 근전도 신호를 이용하는 방법과 발성 없이 근전도 신호만을 이용하는 방법이 있다[비특허문헌 5]. 발성 없는 방법은 무발성 음성인식(Silent Speech Recognition 또는 Mime Speech Recognition)이라고 한다[비특허문헌 5,6]. 인식 대상으로는 대표적으로 영어 단어나 숫자 등 단어(Word)를 단위로 하는 연구들이 있다[비특허문헌 6,9]. 그러나 단어 단위 연구는 데이터베이스에 존재하지 않는 단어를 판별할 가능성이 희박하며, 언어에 의존적이라는 한계를 가진다. 언어의 구조에서 의미의 차이를 가져올 수 있는 소리의 최소 단위인 음소(Phonemes)를 인식 대상으로 하는 연구는 단어 단위 연구의 한계를 해소할 수 있다[비특허문헌 3,5,8]. 여러 음소 단위를 합성하면 단어를 이룰 수 있으며, 모든 단어가 데이터베이스에 있을 필요 없이 음소들만 데이터베이스에 있으면 되기 때문이다.

[0035] 다음으로, 본 발명의 일실시예에 따른 안면근육 표면근전도 신호기반 단모음인식 방법을 도 2를 참조하여 설명한다.

[0036] 본 발명에서는 발성하지 않고 발음에 사용되는 근육만을 움직이게 하여, 안면근육의 표면 근전도 신호만으로 한국어 단모음 'ㅏ', 'ㅑ', 'ㅓ', 'ㅕ', 'ㅗ', 'ㅛ', 'ㅜ', 'ㅠ'의 7개를 인식한다.

[0037] 도 2에서 보는 바와 같이, 본 발명의 일실시예에 따른 안면근육 표면근전도 신호기반 단모음인식 방법은 (a) 트레이닝을 통해 각 모음의 근육 및 특징에 의한 분류기를 생성하는 단계(S10), (b) 각 모음에 대한 최적 특징을 선정하는 단계(S20); (c) 안면근육들의 조합을 최적화하는 단계(S30); 및, (d) 최적화된 안면근육의 조합으로 모음을 인식하는 단계(S40)로 구성된다.

[0038] 먼저, 훈련용 데이터(또는 트레이닝 데이터)로 분류기를 생성한다(S10).

[0039] 트레이닝 단계(S10)에서 표면 근전도 신호는 5개 근육에서 측정되며, 측정된 신호는 특징 추출 알고리즘을 통해 특징값으로 추출된다. 여기서 사용된 특징 추출 알고리즘은 총 5가지로, RMS(Root Mean Square), VAR(Variance), MMAV1(Modified Mean Absolute Value 1), MMAV2(Modified Mean Absolute Value 2), 캡스트럼 계수(Cepstral Coefficients)이다[비특허문헌 11,12]. 5개 근육과 5개의 특징 값으로 변환된 신호들은 25개의 값을 갖는 트레이닝 데이터 그룹으로 생성된다.

[0040] 즉, 사람이 하나의 모음을 발음하면, 이로부터 변환된 신호들로부터 25개의 값(또는 특징값)들이 구해지고, 각각의 값들이 해당 분류기 각각을 트레이닝 시킨다. 따라서 안면근육과 특징의 조합에 의한 25개의 분류기들이

각각 생성된다.

- [0041] 각 분류기는 모음 발음시 측정된 해당 안면근육의 근전도 신호를 해당 특징의 특징추출 알고리즘에 의해 특징값을 추출하면, 추출된 특징값을 입력하여 해당 모음인지 여부를 판단하여 분류한다.
- [0042] 훈련용 데이터는 모음을 발음할 때의 각 근육에서의 근전도 신호 데이터들과 발음한 모음들을 말한다. 예를 들어, 실험자(또는 사람)가 50개의 모음을 각각 발음할 때 얻어진 안면근육의 근전도 신호들(5가지의 근전도 신호들)과, 해당 모음이다. 이 경우, 각 분류기는 50개의 특징값에 의해 트레이닝된다.
- [0043] 다음으로, 각 모음을 정확하게 인식하는 최적의 특징을 선정한다(S20).
- [0044] 생성된 트레이닝 데이터 그룹은 근육별로 인식 정확도가 가장 높은 특징을 사용한다.
- [0045] 단모음마다 근육별로 5개의 특징 값 중에서 가장 높은 인식 정확도를 가진 특징을 선정한다. 예를 들어, ‘ㅏ’ 발음의 1번째 근육에서 특징 값의 인식 정확도가 RMS가 가장 높고, 2번째 근육에서 MMAV1이 가장 높고, 3, 4번째 근육에서 캡스터럼 계수(Cepstral Coefficients)가 가장 높고, 5번째 근육에서 MMAV2가 가장 높으면, ‘ㅏ’ 발음의 각 근육의 특징은 각각 RMS, MMAV1, 캡스터럼 계수(Cepstral Coefficients), 캡스터럼 계수(Cepstral Coefficients), MMAV2가 선정된다.
- [0046] 인식 정확도는 트레이닝 데이터에 의해 발음된 모음을 해당 모음으로 인식하는 확률이다.
- [0047] 이때 분류기는 QDA(Quadratic Discriminant Analysis)와 HMM(Hidden Markov Model)을 사용한다.
- [0048] 다음으로, 각 모음을 정확하게 인식하는 안면근육들의 조합을 최적화한다(S30).
- [0049] 즉, 단모음별로 인식 정확도가 가장 높은 순으로 1개부터 5개까지 근육을 조합하여 단모음의 인식 정확도를 산출하고, 그 중에서 인식 정확도가 가장 높은 근육 조합을 최적화 조합으로 한다. 다수 개의 근육 조합, 즉, 해당 근육에 의한 분류기들의 조합에 의해, 해당 모음을 인식하는 인식 정확도를 산출한다.
- [0050] 이때, 근육의 조합에 따라 해당 분류기들을 논리식으로 조합한다. 바람직하게는, 다수의 분류기를 조합하는 경우, 적어도 50% 이상의 분류기가 해당 모음을 모두 인식하면, 해당 모음으로 인식한다. 예를 들어, 2개 근육에 대한 각각 분류기를 조합하는 경우, 적어도 1개의 분류기에 의해 해당 모음으로 인식되면, 해당 모음을 인식하는 것으로 판단한다. 또한, 5개 근육들 각각의 분류기들의 조합인 경우, 적어도 3개 분류기가 해당 모음을 인식하면 해당 모음을 인식하는 것으로 판단한다.
- [0051] 따라서 근육 조합 최적화는 단모음별로 가장 인식 정확도가 높은 근육 조합을 선택하여 인식 단계에 적용한다.
- [0052] 즉, 5개의 근육 중에서 인식 정확도가 가장 높은 순으로 1개에서 5개까지 조합하여 가장 인식 정확도가 높은 근육 조합을 선정한다. 예를 들어 특징 선정이 끝난 ‘ㅏ’ 발음의 5개의 근육의 인식 정확도가 각각 90%, 85%, 95%, 70%, 75%라면, 근육 조합은 인식 정확도가 높은 순서인 3번, 1번, 2번, 5번, 4번 순으로 조합을 형성한다. 이 때 조합 개수는 1개부터 5개까지로 하며, 근육을 조합하면 각 조합마다 인식 정확도가 계산되고, 이 중에서 가장 높은 근육 조합을 ‘ㅏ’ 발음에 적용한다. 여기서 적용된 특징과 근육 조합이 인식 단계(S40)에서 이용된다.
- [0053] 다음으로, 인식 단계(S40)에서는 새로운 데이터를 받아 트레이닝 단계(앞서 선정한 분류기)와 동일하게 특징 값을 추출하고, 단모음별로 앞서 단계에서 결정된 최적화 조합 근육(해당 조합된 근육들에 대한 분류기들의 조합)을 사용한다.
- [0054] 다음으로, 데이터 수집을 위한 안면근육에 대하여 보다 구체적으로 설명한다.
- [0055] 데이터 수집에 사용된 안면근육은 총 5개로 도 4와 같으며, 말을 할 때 주로 사용되는 근육을 사용하였다. 1채널부터 5채널까지 각각 악이복근 전복, 입꼬리올림근, 대관골근, 교근, 광경근이다. 2, 5채널은 우측, 3, 4채널은 좌측 근육을 사용하였다. 기준전극은 꼭지돌기(Mastoid)를 사용한다.
- [0056] 다음으로, 근전도 신호로부터 특징을 추출하는 특징 추출 알고리즘에 대하여 보다 구체적으로 설명한다.
- [0057] RMS, VAR, MMAV1, MMAV2는 종래 기술에서 한국어 단모음 인식 정확도가 높은 것으로 확인된 특징 추출 알고리즘이다. 위 특징들은 모두 시간 영역(Time Domain)에서 계산된다[비특허문헌 13]. RMS는 일정한 힘과 근수축에 관련되어 있고, VAR은 근전도 신호의 힘을 나타낸다. MMAV1은 MAV(Mean Absolute Value)의 확장형으로 근육의 활동량에 관련되며 가중 윈도우(Weighting Window)를 사용하였다. MMAV2는 MMAV1과 관련되며 연속가중 윈도우

(Continuous Weighting Window)를 사용하였다[비특허문헌 14].

[0058] 캡스트럼(Cepstral Coefficient)은 음성 신호를 분석하기 위해 사용되는 특징으로 주파수 특성을 이용하며, 근전도 신호를 기반으로 한 음성인식에도 적용될 수 있다. 캡스트럼(Cepstral Coefficient)는 다음과 같이 계산된다.

[0059] [수학식 1]

$$c_n = F^{-1} \log |X(f)|$$

[0061] $X(f)$ 는 신호의 주파수 스펙트럼을 나타내며, F^{-1} 는 역푸리에 변환(Inverse Fourier Transform)을 나타낸다. 캡스트럼(Cepstrum)의 계수(Coefficient)는 $n=0$ 일 때의 값을 사용하였다.

[0062] 도 5는 5개의 근육별로 수집된 근전도 신호를 캡스트럼 계수(Cepstral Coefficients)와 RMS 특징값으로 계산한 뒤, PCA(Principal Component Analysis, 주성분 분석)를 통해 5개 근육의 차원을 2차원으로 축소하여 나타낸 특징값의 상관관계 그래프이다.

[0063] 다음으로, 특징값을 입력으로 모음을 분류하는 분류기에 대하여 보다 구체적으로 설명한다.

[0064] QDA(이중판별분석, Quadratic discriminant analysis)는 주로 기계학습과 통계적 분류기로 사용되며, 근전도 신호 분류 알고리즘으로도 사용된다. QDA는 측정된 객체 값들을 이차곡면(Quadric Surface) 위에서 둘 이상의 클래스(Class)로 분류하는 이차 분류기(Quadratic Classifier)중 한가지이다.

[0065] 각 클래스 $k=1,2,\dots,K$ 에 대해 π_k 를 클래스 k 의 사전 확률(Prior Probability), μ_k 를 클래스 k 의 평균 벡터(Mean Vector), Σ_k 를 클래스 k 의 공분산행렬(Covariance Matrix)로 두면, 2차 판별 함수(Quadratic Discriminant Function) δ_k 는 다음과 같다.

[0066] [수학식 2]

$$\delta_k(x) = -\frac{1}{2} \log |\Sigma_k| - \frac{1}{2} (x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k) + \log(\pi_k)$$

[0068] 도 6은 'ㄸ', 'ㄹ', 'ㄴ' 발음을 캡스트럼 계수(Cepstral Coefficients) 특징값으로 변환하여 QDA를 통해 분류된 그래프이다. 실선은 'ㄸ'와 'ㄹ' 발음을, 대시선은 'ㄹ'와 'ㄴ' 발음을, 점선은 'ㄸ'와 'ㄹ' 발음을 QDA로 구분한 선이다.

[0069] HMM(hidden markov model)은 비교적 잘 알려진 통계적 모델이며, 자동음성인식에 주로 사용된다. 사람이 말을 할 때, 음성 신호와 동시에 발생하는 근전도 신호는 같은 문맥적인 정보를 가진다. 따라서 연속적인 근전도 신호는 HMM으로 모델링 될 수 있다.

[0070] 도 7에서 s1~7은 숨겨진 상태를 나타내며, a_{ij} 는 i 에서 j 로의 상태전이 확률이고 b2~6은 관측할 수 있는 결과값의 결과전이 확률이다. 각 단모음별로 HMM을 생성하고 입력 데이터에 대해 관측된 결과로 인식하게 된다.

[0071] 다음으로, 실험을 통한 본 발명의 효과를 구체적으로 설명한다.

[0072] 먼저, 실험 환경을 설명한다.

[0073] 피험자는 조음에 문제가 없는 정상인 12명(남성 10명, 여성 2명)으로 평균 나이 23.25세의 성인이다. 'ㄸ' 발음부터 'ㄴ' 발음까지 7개의 단모음을 차례로 발음하는 것을 1 세트(set)로 하여, 총 100세트를 발음하였다. 50세트는 트레이닝 데이터로, 50세트는 단모음 인식 데이터로 사용되었고, 소리를 내지 않고 근육만을 사용하였다.

[0074] 표면근전도 신호는 근육에 지름 30mm의 양극성 Ag/AgCl 전극을 부착하고, BIOPAC사의 MP150 1대와 BN-EMG(2CH) 3대를 이용하여 취득하였다. 실험에 사용된 장비는 도 8과 같다.

[0075] 다음으로, 실험 결과를 설명한다.

[0076] 실험은 성별 구분을 하지 않았으며, 근육별로 인식 정확도가 가장 높은 특징을 사용하였다. 근육은 가장 인식 정확도가 높은 근육 1개만을 사용한 것(MIX1), 인식 정확도가 높은 순으로 2개(MIX2), 3개(MIX3), 4개(MIX4)를 조합한 것, 모든 근육을 조합한 것(MIX5)을 비교하였다. 근육 조합 최적화는 단모음에 따라 가장 좋은 인식 정확도를 가진 근육 조합이 되도록 하였다.

- [0077] 도 9는 QDA를 사용하여 피험자 3과 피험자 8의 근육에 따른 단모음별 인식 정확도를 나타낸 그래프이다. 그래프는 인식 정확도를 5단계의 범위로 나누어 색상으로 표시하였다. 근육별로는 인식 정확도가 가장 높은 특징이 사용되었다. 피험자 3은 1채널에서 'ㅏ'(66%), 'ㅡ'(62%), 3채널에서 'ㅓ'(88%), 'ㅓ'(100%), 'ㅣ'(72%), 4채널에서 'ㅕ'(100%), 'ㅕ'(68%), 5채널에서 'ㅕ'(100%), 'ㅓ'(100%) 발음의 인식 정확도가 가장 높았고, 피험자 8은 1채널에서 'ㅏ'(54%), 'ㅓ'(80%), 2채널에서 'ㅣ'(100%), 3채널에서 'ㅕ'(100%), 'ㅡ'(52%), 'ㅣ'(100%), 5채널에서 'ㅕ'(96%), 'ㅓ'(50%) 발음의 인식 정확도가 가장 높았다. 특히 피험자 3은 'ㅓ' 발음의 경우 모든 근육의 인식 정확도가 80% 이상이었으며, 피험자 8은 'ㅣ' 발음의 경우 5채널의 근육을 제외한 모든 근육의 인식 정확도가 80% 이상이었다. 피험자의 단모음 발음에 따라 근육별로 인식 정확도가 다른 것을 확인하였고, 이는 단모음 발음에 따른 근육 활성도가 다르다는 것을 의미한다.
- [0078] 도 10은 피험자 전체의 근육별 사용된 특징 횟수를 나타낸다. 단모음별로 50번의 발음 중에서 가장 인식 정확도가 높은 특징이 사용된 횟수를 계수하였으며, 5개 근육에서 각각 계수된 횟수를 특징별로 합하였다. QDA에서 Cepstral Coefficients가 다른 특징들에 비해서 많이 선택되었다. 이는 QDA를 사용할 때 Cepstral Coefficients가 유효한 특징으로 사용될 수 있음을 나타낸다. 하지만 HMM에서는 5개 특징 모두 비슷한 경향을 보였다.
- [0079] 도 11은 피험자 전체의 근육 조합별 인식 정확도와 근육 최적화에 따른 인식 정확도를 나타낸 그래프이다. 근육 조합은 인식 정확도가 높은 순으로 3개를 조합하는 경우가 'ㅡ' 발음에서 가장 높았으며, 나머지 발음에서는 2개를 조합하는 경우가 가장 높았다.
- [0080] QDA에서는 각 단모음별로 근육 조합을 최적화하는 방법이 인식 정확도가 전체 평균 85.7%로 가장 높았으며, 이어서 2개 조합이 평균 84.2%, 3개 조합이 평균 82.7%를 보였다. HMM에서는 근육 조합 최적화 방법이 평균 75.1%, 2개 조합이 평균 73.3%, 3개 조합이 72.2%를 보였다.
- [0081] 도 12의 표는 전체 피험자의 한국어 단모음 인식 혼동행렬(Confusion Matrix)을 나타낸 것이다. 각 행은 실제로 발음한 단모음이고, 각 열은 단모음별로 인식된 개수이다. 검은색으로 칠해진 곳은 맞게 인식된 것을 나타내고, 회색으로 칠해진 곳은 인식 오류율이 10% 이상인 것을 나타낸다. 'ㅏ' 발음과 'ㅕ' 발음이 서로 잘못 인식되는 경우가 많았고, HMM의 경우 'ㅡ' 발음과 'ㅣ' 발음은 'ㅏ' 발음으로 잘못 인식되는 경우가 많았다.
- [0082] 또한, 도 11에서 'ㅏ', 'ㅕ', 'ㅡ', 'ㅣ' 발음이 다른 발음에 비해 인식 정확도가 다소 떨어진다. 이에 대한 오차의 원인은 도 5를 보면 'ㅏ'와 'ㅕ', 'ㅡ'와 'ㅣ'의 특징 값이 상당히 겹치게 되는 것에서 찾을 수 있다.
- [0083] 도 13은 제안방법과 다른 연구들의 평균 인식 정확도를 비교한 그래프이다. Kumar 등이 제안한 방법은 ANN(Artificial Neural Network)으로 영어 모음 5개를 분류하였으며, 평균 인식 정확도는 88.9%이다[비특허문헌 3]. 이는 본 발명에서 제안한 방법과 큰 차이를 보이지 않으며, 본 발명에서 제안한 방법은 12명의 피험자와 7개의 단모음을, Kumar 등이 제안한 방법은 3명의 피험자와 5개의 모음을 분류한 것이다. 또한 발성이 있는 근전도 신호이기 때문에 정확도가 다소 높을 수 있다. Manabe 등이 제안한 방법은 Multi-stream HMM으로 일본어 모음 5개를 분류하였다[비특허문헌 6]. 평균 인식 정확도는 63.7%이다. Srisuwan 등이 제안한 방법은 Neuron Network로 태국어 5개의 톤(Tone)을 분류하였다[비특허문헌 14]. 평균 인식 정확도는 80%이다. Lopez-Larraz 등이 본 발명에 따른 방법은 Decision Tree와 AdaBoost로 영어 모음 5개를 분류하였고, 평균 80.2%의 인식 정확도를 보였다[비특허문헌 15].
- [0084] 본 발명에 따른 한국어 단모음 7개를 분류하는 방법은 QDA에서 85.7%로 유효한 인식 정확도를 보이지만, HMM은 75.1%로 다소 떨어지는 인식 정확도를 보였다.
- [0085] 본 발명에서는 발성 없이 근전도 신호만으로 한국어 단모음 7개를 근육 최적화를 통해 인식하는 방법을 설명하였다. 단모음에 따라 근육 활성도가 다르다는 것을 이용하여 근육 조합을 최적화하여 단모음을 인식하였으며, 다른 시간 영역의 특징보다는 캡스트럼 계수(Cepstral Coefficients)가 효과적인 것을 확인하였다.
- [0086] 이상, 본 발명자에 의해서 이루어진 발명을 실시 예에 따라 구체적으로 설명하였지만, 본 발명은 실시 예에 한정되는 것은 아니고, 그 요지를 이탈하지 않는 범위에서 여러 가지로 변경 가능한 것은 물론이다.
- [0087]

부호의 설명

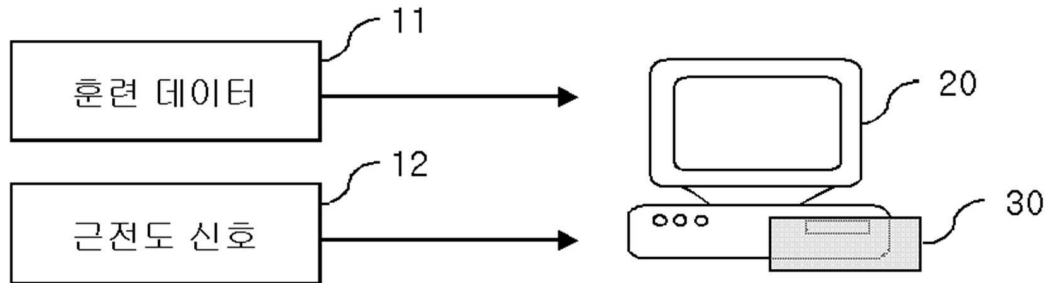
- [0088] 10 : 훈련데이터 12 : EMG 신호

20 : 컴퓨터 단말

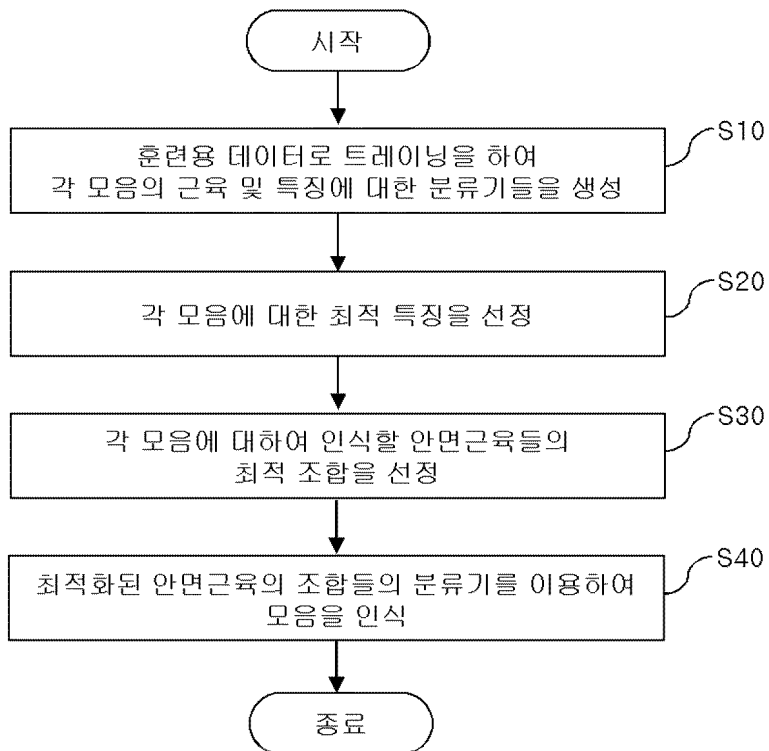
30 : 사용자 인식 장치

도면

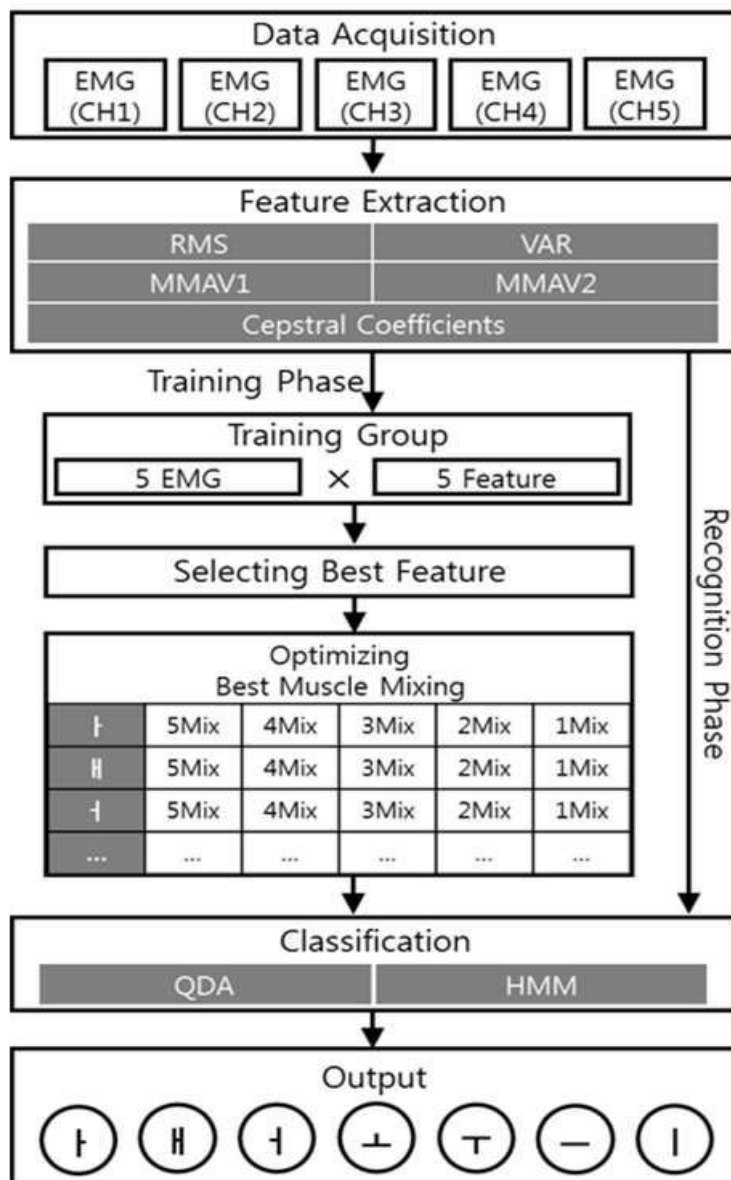
도면1



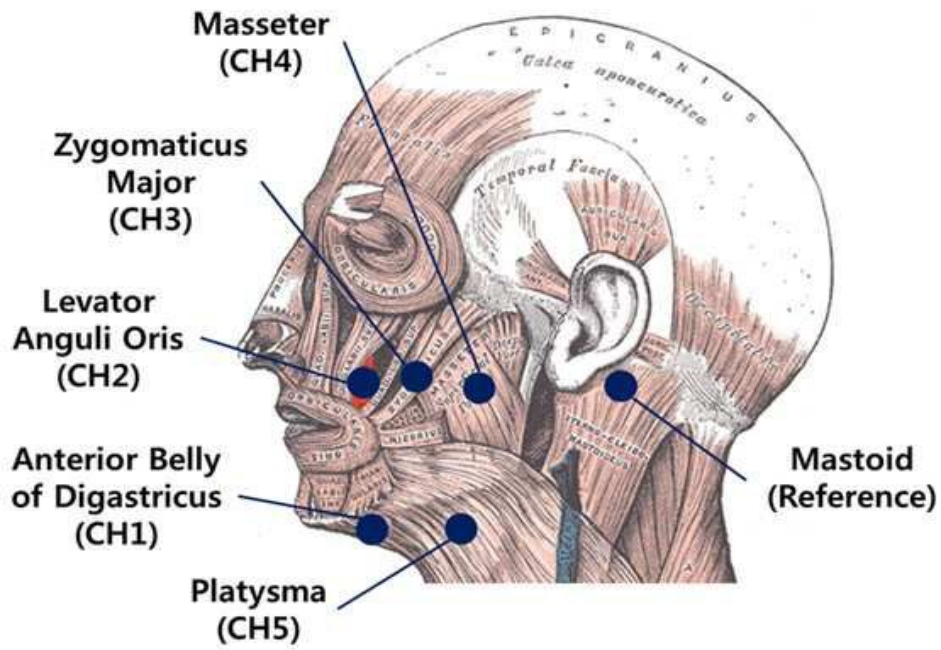
도면2



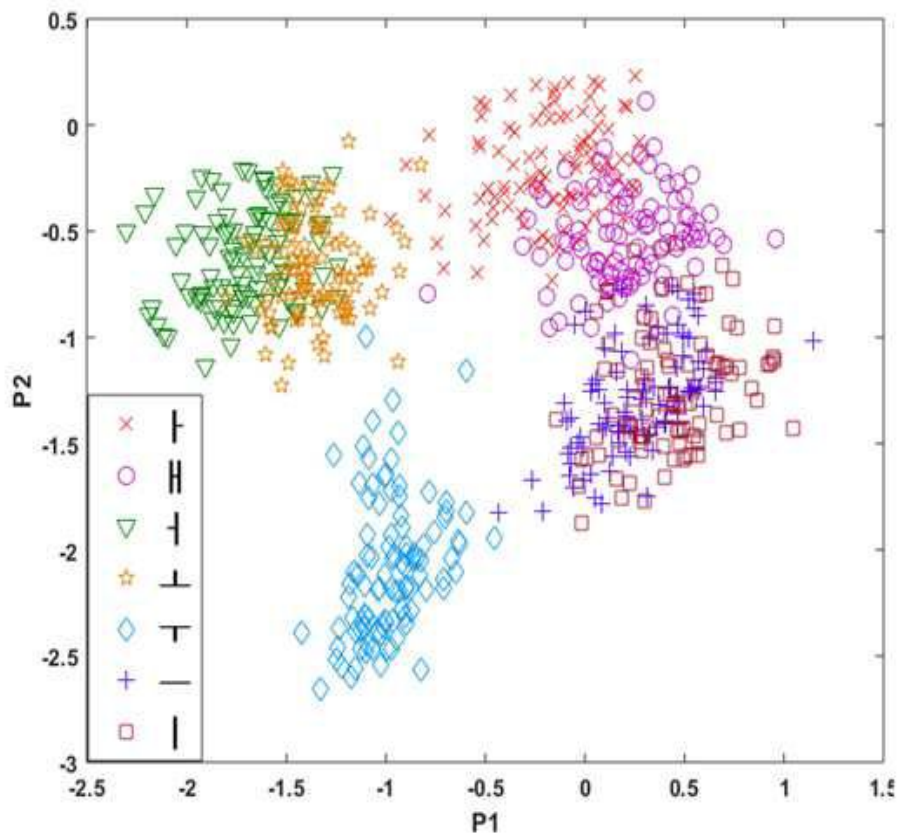
도면3



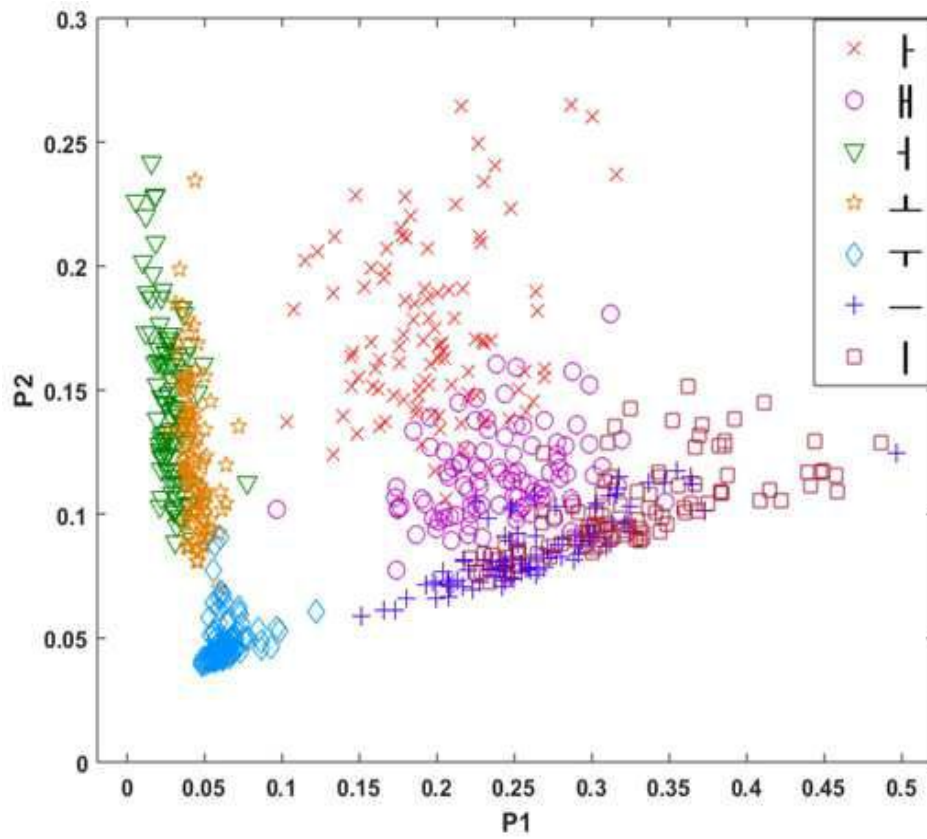
도면4



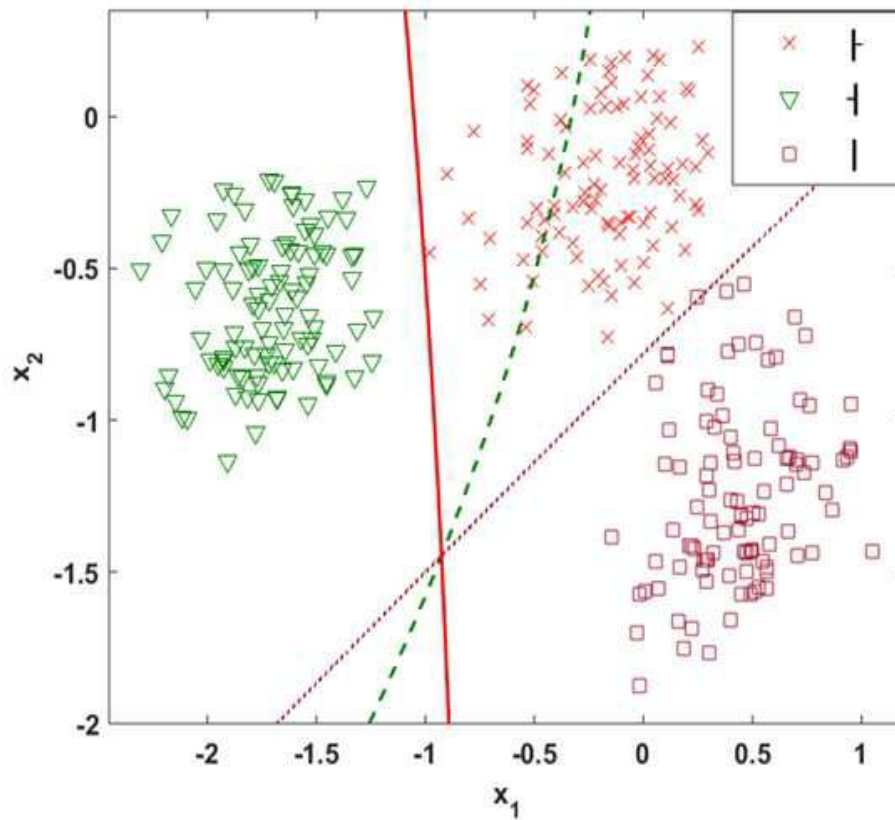
도면5a



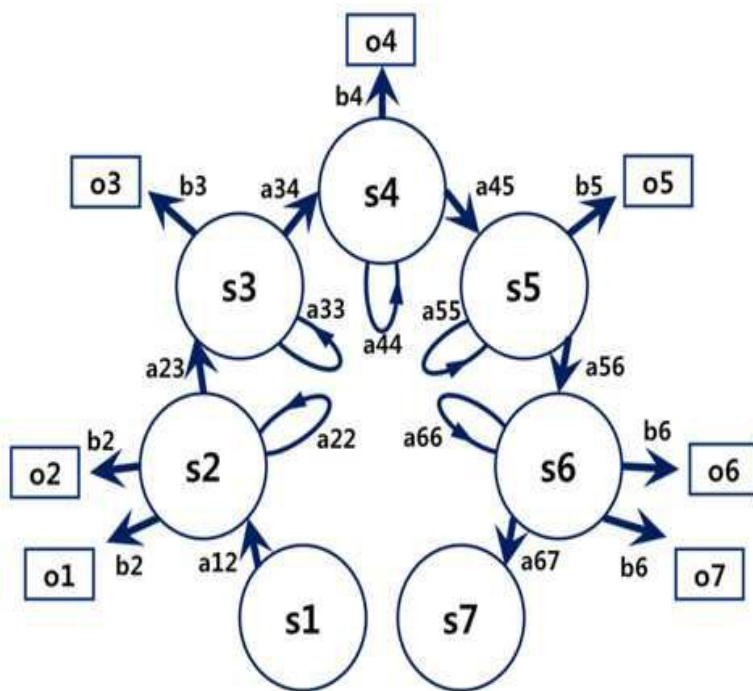
도면5b



도면6



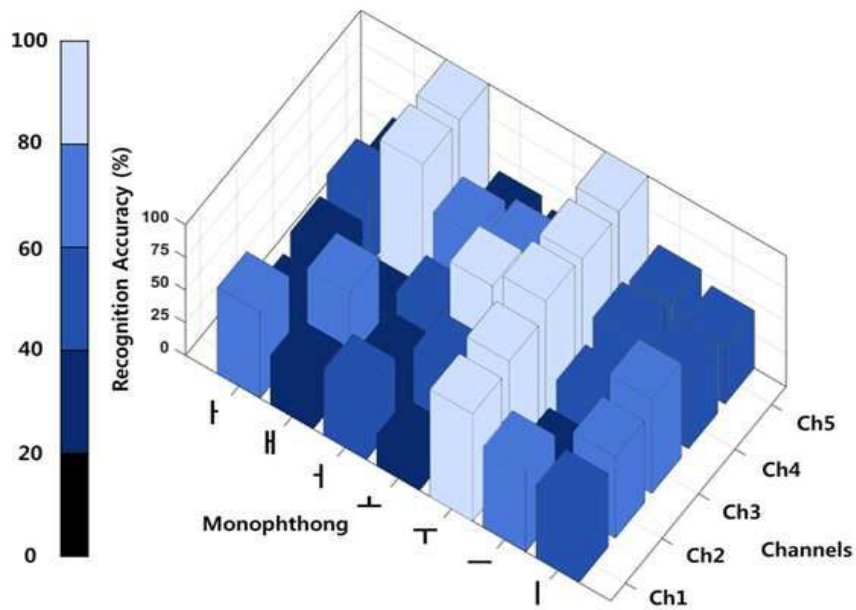
도면7



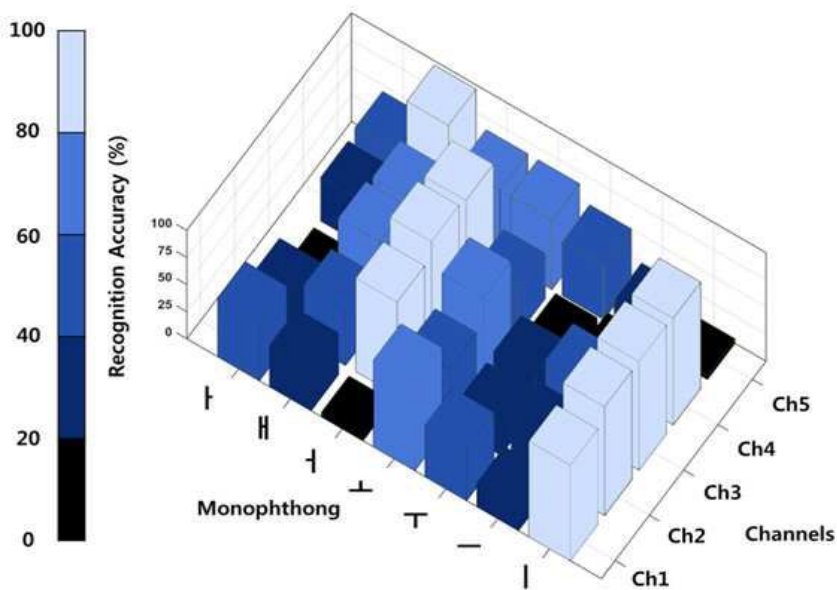
도면8



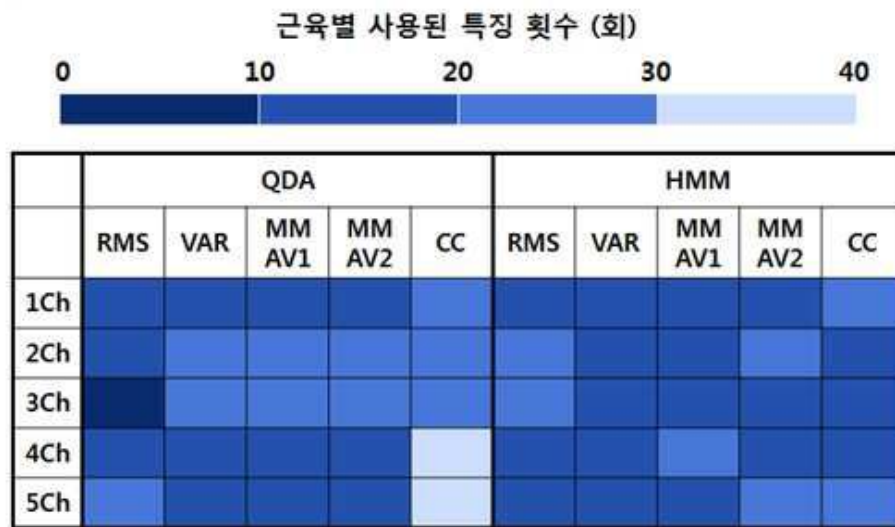
도면9a



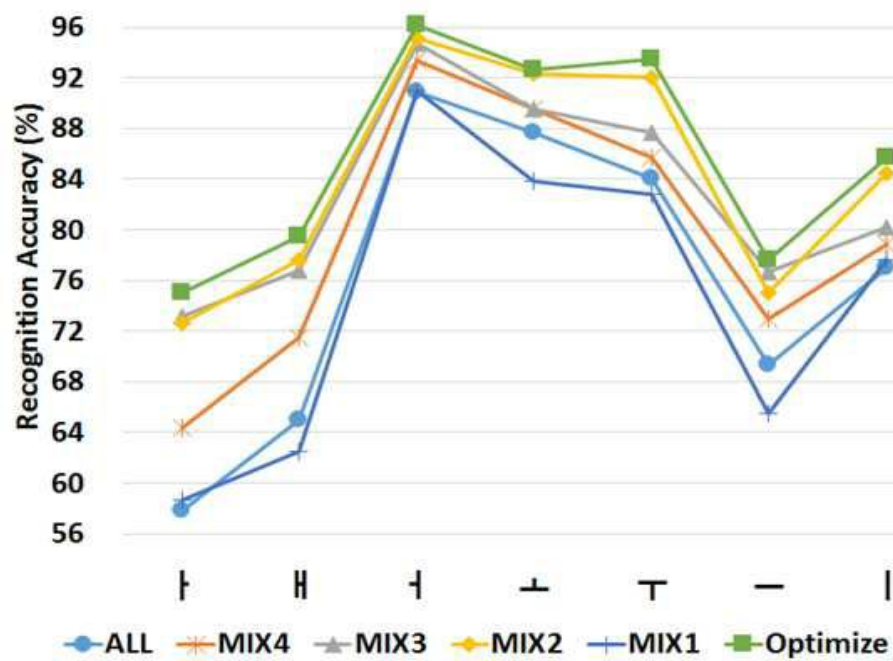
도면9b



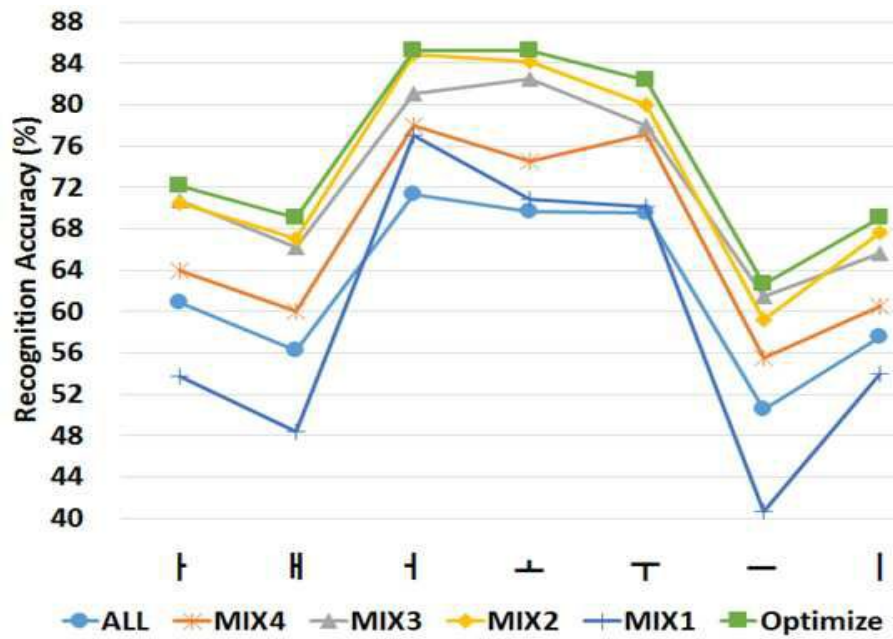
도면10



도면11a



도면11b



도면12

QDA		Recognized Monophthong							Accuracy (%)
		ㅏ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ		
Phonated Monophthong	ㅏ	450	37	59	35	10	5	4	75.0
	ㅓ	54	477	26	11	10	16	6	79.5
	ㅗ	2	4	577	13	4	0	0	96.2
	ㅜ	11	5	15	556	13	0	0	92.7
	ㅡ	12	3	5	10	561	6	3	93.5
	ㅣ	32	25	20	13	11	466	33	77.7
	ㅣ	38	25	0	1	6	16	514	85.7
HMM		Recognized Monophthong							Accuracy (%)
		ㅏ	ㅓ	ㅗ	ㅜ	ㅡ	ㅣ		
Phonated Monophthong	ㅏ	433	86	32	23	13	3	10	72.2
	ㅓ	89	414	48	20	15	11	3	69.0
	ㅗ	19	35	511	34	1	0	0	85.2
	ㅜ	17	36	19	511	13	4	0	85.2
	ㅡ	17	19	12	51	494	5	2	82.3
	ㅣ	102	53	13	12	24	376	20	62.7
	ㅣ	125	24	2	7	9	19	414	69.0

도면13

