

Original Article  
Neuroscience



# Assessment of Dysarthria Using One-Word Speech Recognition with Hidden Markov Models

Seung Hak Lee <sup>1,2\*</sup> Minje Kim <sup>3</sup> Han Gil Seo <sup>1</sup> Byung-Mo Oh <sup>1</sup>  
Gangpyo Lee <sup>2</sup> and Ja-Ho Leigh <sup>2,4,5†</sup>

<sup>1</sup>Department of Rehabilitation Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

<sup>2</sup>Department of Rehabilitation Medicine, Incheon Workers' Compensation Hospital, Incheon, Korea

<sup>3</sup>Department of Intelligent Systems Engineering, Indiana University, Bloomington, IN, USA

<sup>4</sup>Department of Rehabilitation Medicine, Incheon St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Incheon, Korea

<sup>5</sup>Institute of Public Health and Medical Service, Seoul National University Hospital, Seoul, Korea



Received: Jan 5, 2019

Accepted: Mar 24, 2019

**Address for Correspondence:**

Ja-Ho Leigh, MD, MS

Department of Rehabilitation Medicine, Seoul National University Hospital, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Republic of Korea; Department of Rehabilitation Medicine, Incheon Workers' Compensation Hospital, 226 munemi-ro, Bupyeong-gu, Incheon 21417, Republic of Korea.  
E-mail: jaho.leigh@gmail.com

\*Current address: Department of Rehabilitation Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

†Current address: Department of Rehabilitation Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

© 2019 The Korean Academy of Medical Sciences.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**ORCID iDs**

Seung Hak Lee 

<https://orcid.org/0000-0002-3017-8497>

<https://jkms.org>

## ABSTRACT

**Background:** The gold standard in dysarthria assessment involves subjective analysis by a speech–language pathologist (SLP). We aimed to investigate the feasibility of dysarthria assessment using automatic speech recognition.

**Methods:** We developed an automatic speech recognition based software to assess dysarthria severity using hidden Markov models (HMMs). Word-specific HMMs were trained using the utterances from one hundred healthy individuals. Twenty-eight patients with dysarthria caused by neurological disorders, including stroke, traumatic brain injury, and Parkinson's disease were participated and their utterances were recorded. The utterances of 37 words from the Assessment of Phonology and Articulation for Children test were recorded in a quiet control booth in both groups. Patients were asked to repeat the recordings for evaluating the test–retest reliability. Patients' utterances were evaluated by two experienced SLPs, and the consonant production accuracy was calculated as a measure of dysarthria severity. The trained HMMs were also employed to evaluate the patients' utterances by calculating the averaged log likelihood (aLL) as the fitness of the spoken word to the word-specific HMM.






**Results:** The consonant production accuracy reported by the SLPs strongly correlated ( $r = 0.808$ ) with the aLL, and the aLL showed excellent test–retest reliability (intraclass correlation coefficient, 0.964).

**Conclusion:** This leads to the conclusion that dysarthria assessment using a one-word speech recognition system based on word-specific HMMs is feasible in neurological disorders.

**Keywords:** Automatic Speech Recognition; Dysarthria; Hidden Markov Models

## INTRODUCTION

Dysarthria, a neuromotor speech disorder, is a common symptom of various neurological disorders, including stroke, parkinsonism, and traumatic brain injury.<sup>1</sup> With regard to stroke, the prevalence of dysarthria in acute first stroke is 41.5%–53%, and it plays an important role in determining stroke severity.<sup>2</sup> Dysarthria has a profound effect on patient function and severely lowers the quality of life.<sup>3</sup> Different forms of dysarthria exist; each of which has

Minje Kim   
<https://orcid.org/0000-0003-3513-8328>  
 Han Gil Seo   
<https://orcid.org/0000-0001-6904-7542>  
 Byung-Mo Oh   
<https://orcid.org/0000-0001-9353-7541>  
 Gangpyo Lee   
<https://orcid.org/0000-0002-6381-9211>  
 Ja-Ho Leigh   
<https://orcid.org/0000-0003-0465-6392>

### Funding

This research was supported by a grant of the Research Center for Innovation in Medical Rehabilitation, funded by the Korea Workers' Compensation and Welfare Service in 2016.

### Disclosure

The authors have no potential conflicts of interest to disclose.

### Author Contributions

Conceptualization: Lee SH, Kim M, Seo HG, Leigh JH. Data curation: Lee SH, Kim M. Formal analysis: Lee SH, Kim M, Seo HG, Leigh JH, Oh BM. Investigation: Lee SH, Kim M, Seo HG, Leigh JH, Oh BM. Methodology: Lee SH, Kim M, Seo HG, Leigh JH. Software: Kim M. Validation: Kim M. Writing - original draft: Lee SH, Kim M.

different speech characteristics and is related to the site of lesion and degree of neurological damage.<sup>4,5</sup> Clinically, dysarthria assessment is performed by speech–language pathologists (SLPs), who subjectively measure articulation and speech intelligibility.<sup>6</sup> Dysarthria assessment by SLPs requires substantial effort, and the severity of dysarthria is reported as mild, moderate, or severe.<sup>7</sup> Therefore, if there is a tool that can objectively and easily measure the severity of dysarthria, it may be clinically useful to evaluate the disorder and confirm the effect of treatment.

Automatic speech recognition (ASR) is the process by which a computer can recognize spoken language or utterances. Rapid advances in ASR technology have led to the widespread use of ASR systems in various devices, such as smartphones and smart home devices, with a goal of providing an automated assistant system that can transcribe spoken language as accurately as possible. Furthermore, ASR systems specialized for dysarthria have improved ASR performance.<sup>7</sup> Several ASR models based on hidden Markov models (HMMs) or artificial neural networks have been utilized.<sup>8,9</sup> Our hypothesis is that quantitative and objective assessment of dysarthria is possible using these ASR algorithms. For example, the fitness of a spoken keyword as assessed using a word-specific HMM may be correlated with the subjective measurement of dysarthria severity by SLPs. The purpose of the present study was to investigate the feasibility of assessing the severity of dysarthria in patients with neurological diseases in a simple and quantitative way using an HMM-based ASR system.

## METHODS

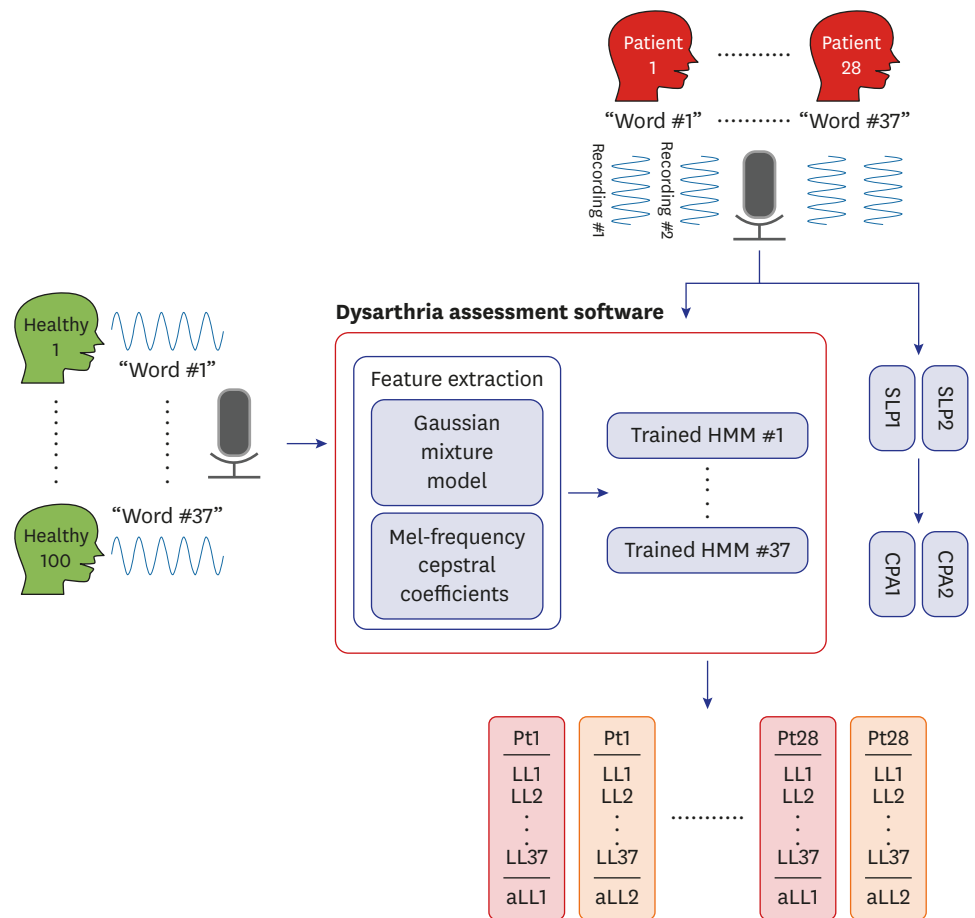
### Participants and study design

Healthy individuals and patients with dysarthria were recruited. One hundred healthy individuals were used to develop word-specific HMMs; thus, individuals who had experienced any type of neurological disease or laryngeal problems (e.g., vocal cord palsy) or had a strong provincial accent were excluded. Twenty-eight patients with various neurological disorders who complained of discomfort in articulation and had speech dysarthria on neurological examination were included. Patients with severe cognitive impairment, aphasia, or pre-morbid strong provincial accent were excluded (**Table 1**). There was one patient with Parkinson's disease who had definite dysarthria without severe hypophonia. All participants were instructed to speak 37 words from the Assessment of Phonology and Articulation for Children (APAC) test, which is a normalized speech assessment tool for the Korean language.<sup>10</sup> The words were displayed on a tablet PC, and the utterances were recorded in a quiet control booth (noise level  $\leq$  40 dB). For the patient group, the recording session was repeated to calculate the test-retest reliability for word-specific HMMs. The patients'

**Table 1.** Clinical characteristics of participants

Parameters	Healthy control	Patients with dysarthria
No.	100	28
Age, yr, mean $\pm$ SD	44.6 $\pm$ 12.9	53.8 $\pm$ 9.5
Gender, men/women	58/42	20/8
Symptom duration, mon, mean $\pm$ SD	-	20.4 $\pm$ 28.3
Ischemic stroke	-	9
Hemorrhagic stroke	-	15
Traumatic brain injury	-	3
Parkinson's disease	-	1

SD = standard deviation.



**Fig. 1.** Study design and word-specific HMM training. Separate HMMs were generated for each of the 37 words in the Assessment of Phonology and Articulation for Children test and HMMs were trained through recordings from 100 healthy individuals. Recordings from 28 patients were applied to the developed dysarthria assessment software to yield aLL. Subsequently, two SLP listen to the recordings and calculate the CPA. Two repeated recordings in the patient group were applied to the software and obtain aLL1 and aLL2 for test-retest reliability. HMM = hidden Markov model, aLL = average log likelihood, CPA = consonant production accuracy.

recorded utterances were blindly evaluated by two SLPs and then assessed by the dysarthria assessment software using word-specific HMMs (Fig. 1).

### Speech examination by SLPs

Two experienced SLPs listened to the recorded utterances by patients and evaluated the consonant production accuracy (CPA). The CPA is defined as the percentage of correct consonant production and is the only quantitative value provided in the APAC test.<sup>11</sup> They listened to repeated recordings from the patient and judged how precisely the consonants were produced in the patient's vocalizations. The two SLPs did not know each other's evaluation results. CPA1 was measured by one SLP and CPA2 by the other.

### Word-specific HMMs

A total of 37 HMM-based ASR systems were trained, one for each of the 37 keywords, using the utterances from the healthy group. The number of hidden states for each word-specific HMM was based on the number of phonemes, and each state was represented by a Gaussian mixture model (GMM) with five modalities. We used standard Mel-frequency cepstral coefficients (MFCCs) and their delta and delta–delta; thus, the GMM approximated a set

of MFCC-related 39-dimensional feature vectors belonging to the same hidden state. Once the training procedure was completed, the patients' utterances of each word were evaluated using the corresponding HMM, and the log likelihood (LL) was calculated as a measure of fitness. The average of the 37 LL values (average log likelihood [aLL]) was considered to reflect dysarthria severity. The aLL obtained by applying the first recording set was regarded as aLL1 and that obtained by applying the second recording set was regarded as aLL2 (Fig. 1).

### Statistical analysis

The Pearson correlation coefficient was used to evaluate the relationship between the aLL and CPA values. Interrater reliability of SLPs and test–retest reliability of dysarthria assessment using HMMs were assessed using the intraclass correlation coefficient. A Bland-Altman plot was used to evaluate the agreement between SLPs.

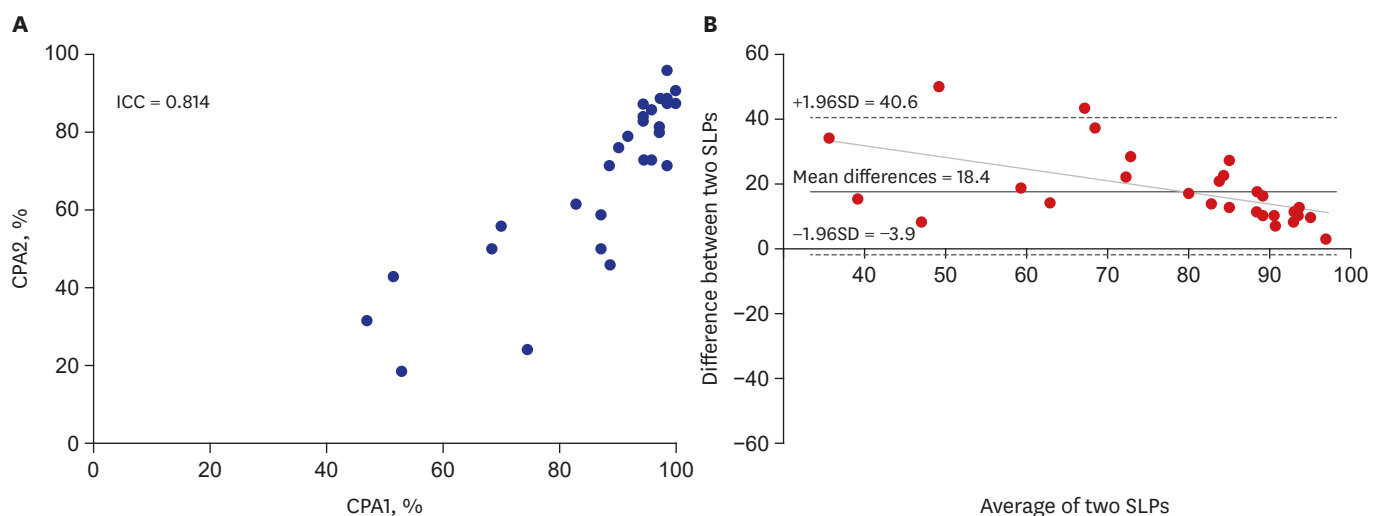
### Ethics statement

The study was approved by the Institutional Review Board (IRB) at Seoul National University Hospital (IRB No. 1603-095-750). All participants provided written informed consent.

## RESULTS

### Assessment of dysarthria by SLPs

The intraclass correlation coefficient between CPA1 and CPA2 was 0.814 (95% confidence interval [CI], 0.779–0.953;  $P < 0.001$ ). However, the Bland-Altman plot showed remarkable biases between the two measurements (Fig. 2). The mean difference (fixed bias) between CPA1 and CPA2 (CPA1–CPA2) was 18.4%. The limits of agreement determined by the 95% CI were –3.9 to 40.6%, suggesting that SLP1 tended to measure CPA at a high level. A proportional bias was also observed, which was more pronounced when the patient's dysarthria was severe.



**Fig. 2.** CPA from two different SLPs. **(A)** The scatter plot for the CPA value of each patient with both measurements from both SLPs **(B)** Bland-Altman plot for the CPA value of each patient with measurements from both SLPs. Mean differences are indicated by the solid line and 95% limits of agreement (mean differences  $\pm$  1.96 standard deviation of the difference) are shown by the dashed line. The dotted line shows regression lines for proportional biases. CPA = consonant production accuracy, SLPs = speech-language pathologists.

### Assessment of dysarthria using HMMs

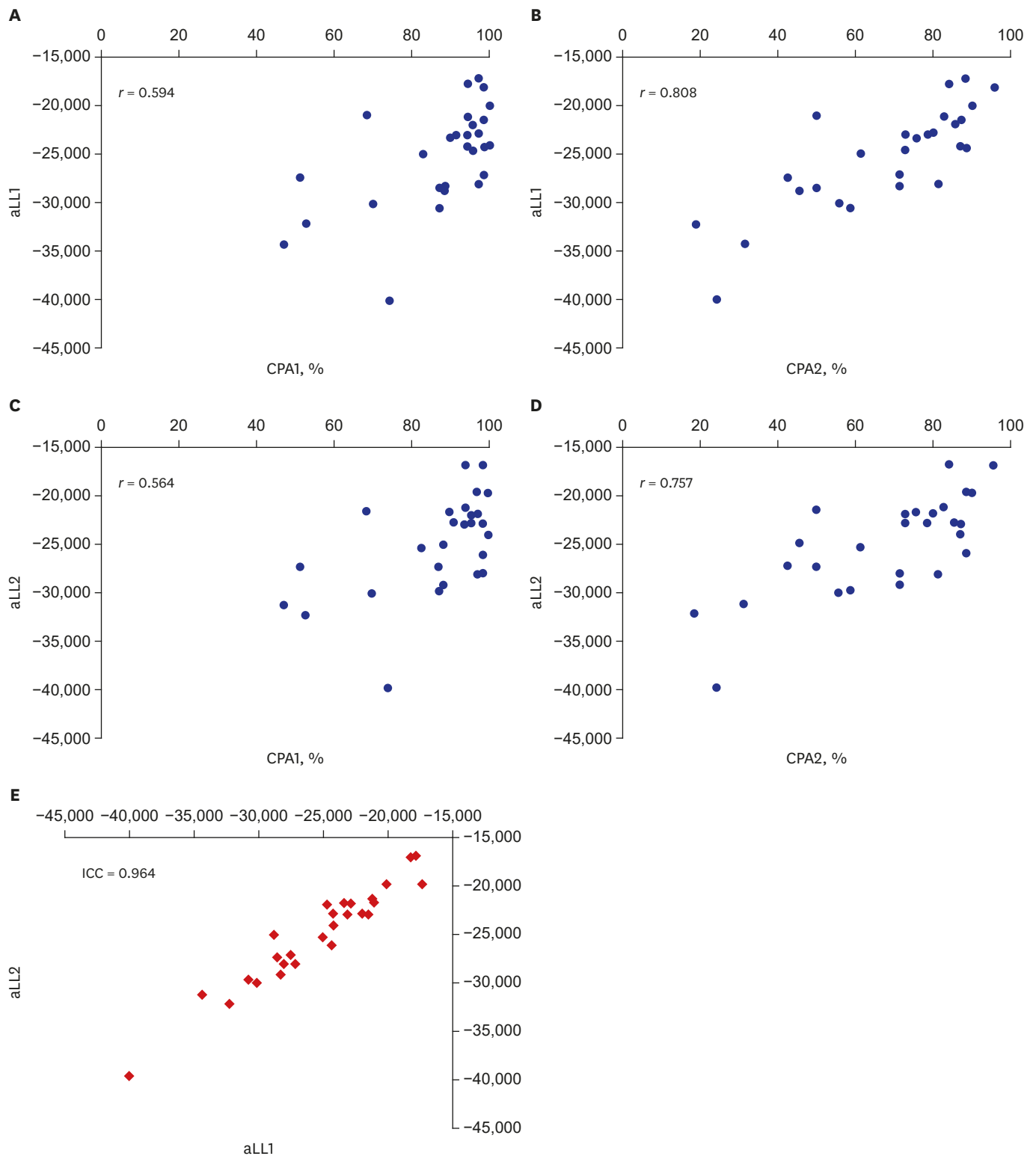
The Pearson correlation coefficients were 0.594 ( $P < 0.01$ ) between CPA1 and aLL1, 0.564 ( $P < 0.01$ ) between CPA1 and aLL2, 0.808 ( $P < 0.01$ ) between CPA2 and aLL1, and 0.757 ( $P < 0.01$ ) between CPA2 and aLL2 (Fig. 3). CPA2 (measured by SLP2) showed a stronger correlation with aLL than CPA1. The intraclass correlation coefficient between aLL1 and aLL2 was 0.964 (95% CI, 0.923–0.983;  $P < 0.001$ ), which means that the test–retest reliability of dysarthria assessment using HMMs was excellent. The correlation coefficients between the word-specific LL and CPA1 ranged from 0.322 to 0.838 (Supplementary Table 1).

## DISCUSSION

In this study, we adopted ASR for the assessment of dysarthria in patients with neurological disorders and compared it with conventional speech evaluation by SLPs. The results showed that the numeric values produced by the custom-made word-specific HMMs correlated with the subjectively measured severity by the SLP. Moreover, assessment of dysarthria using HMMs also showed excellent test–retest reliability.

Our study results demonstrated the subjectivity of SLPs in dysarthria assessment, which was one of our study motivations. Although there was a strong inter-rater reliability between the two SLPs, there was a significant bias as indicated by the Bland-Altman plot. The CPA value measured by SLP1 was higher than that measured by SLP2 in all patients. There has been no study on the reliability of APAC testing in Korean language. However, both intra-rater and inter-rater reliability of auditory-perceptual evaluation of dysarthria fluctuates greatly from study to study.<sup>6,12</sup> This is consistent with our clinical experience and justifies the need to develop quantitative and objective assessment technology, such as an ASR system, for dysarthria. There have been several attempts to assess dysarthria using an acoustic analysis with computer devices with an acceptable reliability.<sup>13,14</sup> However, it has not been widely used in clinics because it is still tedious, complicated, and time-consuming. Our simple approach is seemingly different from the former methodologies because it is not based on the physical parameters from phonological measurement but based on statistical algorithms.

A previous study demonstrated that the ASR recognition rate for dysarthric speech could be predicted by a speech consistency score; however, this measure has not been validated against the gold standard SLP evaluation.<sup>15</sup> In the present study, ASR was adopted for dysarthria assessment in patients with neurological disorders and was compared to conventional speech evaluation by experienced SLPs. Our results showed that the objective fitness measures (aLL1 and aLL2) correlate with the perceptual measures of SLPs (CPA) in terms of Pearson correlations between CPA and aLL, which varied from 0.564 to 0.808. An interesting observation is that the correlation values vary depending on the choice of CPAs and not on aLLs. For example, the difference between Fig. 3A and C is only 0.03, where CPA1 is compared against aLL1 and aLL2, while Fig. 3A and B are quite different by 0.214, where CPA1 and CPA2 are compared against aLL1. One of the reason of this phenomenon is proportional bias among SLPs. SLP2 tends to give worse CPA scores to patients with severe dysarthria. This may have affected the spectrum of severity of dysarthria and the higher Pearson correlations with aLL scores. It suggests that the subjective scores from different SLPs could vary more than the word-specific HMMs tested on different dysarthric speech signals. It also suggests that the proposed HMM-based objective metrics at least has a higher



**Fig. 3.** Comparison between CPA values measured by SLPs and aLL values calculated by word-specific HMMs. **(A-D)** The scatter plot comparing CPA values from SLPs and aLL values from word-specific HMMs in repetition. **(E)** The scatter plot of the aLL value obtained by applying the repeated recordings to the HMM. CPA = consonant production accuracy, SLPs = speech-language pathologists, aLL = average log likelihood, HMM = hidden Markov model.

test–retest reliability than the CPAs from different SLPs, although we cannot claim the overall reliability of the proposed method out of this observation.

Dysarthria assessment using ASR has several advantages. First, the assessment is not dependent on an SLP, and is therefore, objective, quantitative, and low-cost. Second, ASR can be implanted in mobile devices, enabling self-assessment and the use of telemedicine, which has recently become popular in stroke care.<sup>16</sup> Although internet-based telerehabilitation evaluation systems have been proposed for dysarthria, these are based on videoconferencing with an assigned SLP.<sup>17,18</sup> Using ASR, patients could simply speak keywords into their mobile device, and a prompt assessment can be transferred to health facilities for further analysis. However, with the results of this study, it cannot be justified that the ASR assessment can substitute the conventional evaluation by SLPs. Because the aLL is a numeric value that represents the severity of dysarthria, it does not reflect the various clinical features of dysarthria. In addition, detailed assessment by SLPs from a therapeutic viewpoint is crucial to establishing a treatment plan for dysarthria. Accordingly, our methodology could be used for only the screening or follow-up of dysarthric patients and further investigation is necessary.

The present study has several limitations. First, the healthy individuals used to train the HMMs were likely not the ideal references. Speech characteristics are highly dependent on age and gender. Furthermore, the voices of the patients with dysarthria after stroke are similar to those of healthy aging speakers.<sup>19</sup> Thus, HMMs trained using healthy age-matched individuals may render better dysarthria assessment performance. Second, we had only two SLP assessment results, and the comparison between these and the ASR system is limited. Third, the average of log likelihood obtained from 37 word-specific HMMs can also be criticized as an indicator of dysarthria severity. HMM produces a probability value, which is extremely small and is converted to a logarithmic scale. Therefore, a direct comparison between the aLL value and the CPA measured as a percentage is difficult, and the agreement between the two methods cannot be investigated.

From the technical viewpoint of speech recognition, there is still room for improvement. First, instead of designing the HMM to have the same number of hidden states for the phonemes in the target word, a left-to-right HMM model can help improve the training part.<sup>20</sup> More importantly, an HMM-based recognition system that is trained from clean speech signals is known to be sensitive to the artifacts that a real-world speech signal can be exposed to, e.g. reverberation, additive interfering signals, band-pass filtering, etc. For example, a dysarthric utterance recorded in a quiet room can get a better score than a normal but noisy utterance. In this study, we get around this problem by recording the dysarthric utterances in the same quiet booth used for collecting normal speech, but this convenient assumption does not always hold true in real-world recordings. Therefore, a practical system must employ the speech enhancement methods to pre-process the real-world recordings before feeding them to the ASR module. Another possible reason of the discrepancy between the subjective APAC scores and the log likelihood values of HMMs could be the fact that HMM takes both consonants and vowels into account, whereas the criteria the APAC test uses are based only on consonants.

Although this is a factor that might degrade the correlation score reported in the Result section, it could also mean that the proposed HMM-based system can be more promising than the APAC test as it evaluates not only consonants but also vowels. The present results suggest that dysarthria evaluation using a thirty seven-words speech recognition system

based on word-specific HMMs is feasible in neurological disorders. However, there are many limitations to substitute conventional assessment by SLPs with this system and further upgrades are needed for it to be used in clinical practice.

## ACKNOWLEDGMENTS

The authors thank Professor Seung Ha Oh of the Department of Otolaryngology, Seoul National University Hospital, for providing the recording booth.

## SUPPLEMENTARY MATERIAL

### Supplementary Table 1

Word-specific data

[Click here to view](#)

## REFERENCES

1. Duffy JR. *Motor Speech Disorders*. 3rd ed. St. Louis, MO: Elsevier Health Sciences; 2013.
2. Mackenzie C. Dysarthria in stroke: a narrative review of its description and the outcome of intervention. *Int J Speech-Language Pathol* 2011;13(2):125-36.  
[PUBMED](#) | [CROSSREF](#)
3. Ali M, Lyden P, Brady M; VISTA Collaboration. Aphasia and dysarthria in acute stroke: Recovery and functional outcome. *Int J Stroke* 2015;10(3):400-6.  
[PUBMED](#) | [CROSSREF](#)
4. Rampello L, Rampello L, Patti F, Zappia M. When the word doesn't come out: A synthetic overview of dysarthria. *J Neurol Sci* 2016;369:354-60.  
[PUBMED](#) | [CROSSREF](#)
5. Urban PP, Rolke R, Wicht S, Keilmann A, Stoeter P, Hopf HC, et al. Left-hemispheric dominance for articulation: a prospective study on acute ischaemic dysarthria at different localizations. *Brain* 2006;129(Pt 3):767-77.  
[PUBMED](#) | [CROSSREF](#)
6. Wannberg P, Schalling E, Hartelius L. Perceptual assessment of dysarthria: comparison of a general and a detailed assessment protocol. *Logoped Phoniatr Vocol* 2016;41(4):159-67.  
[PUBMED](#) | [CROSSREF](#)
7. Young V, Mihailidis A. Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: a literature review. *Assist Technol* 2010;22(2):99-112.  
[PUBMED](#) | [CROSSREF](#)
8. Hasegawa-Johnson M, Gunderson J, Penman A, Huang T. HMM-based and SVM-based recognition of the speech of talkers with spastic dysarthria. In: *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*. 2006, 3.
9. Hinton G, Deng L, Yu D, Dahl G, Mohamed A, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* 2012;29(6):82-97.  
[CROSSREF](#)
10. McLeod S, Verdon S. A review of 30 speech assessments in 19 languages other than English. *Am J Speech Lang Pathol* 2014;23(4):708-23.  
[PUBMED](#) | [CROSSREF](#)
11. Kim MJ, Pae SY, Park CI. *Assessment of Phonology for Children (APAC)*. Incheon: Human Brain Research & Consulting Co.; 2007.



12. Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res* 1993;36(1):21-40.  
[PUBMED](#) | [CROSSREF](#)
13. Kent RD, Vorperian HK, Duffy JR. Reliability of the multi-dimensional voice program for the analysis of voice samples of subjects with dysarthria. *Am J Speech Lang Pathol* 1999;8(2):129-36.  
[CROSSREF](#)
14. Kim YH, Kim WH, Kim HG. A study on acoustic characteristics of dysarthria in relation to the underlying etiology. *J Korean Acad Rehabil Med* 1994;18(4):12.
15. Kayasith P, Theeramunkong T, Thubthong N. Recognition rate prediction for dysarthric speech disorder via speech consistency score. In: Yang Q, Webb G, editors. *Pacific Rim International Conference on Artificial Intelligence*. Berlin, Heidelberg: Springer; 2006, 885-9.
16. Schwamm LH, Holloway RG, Amarenco P, Audebert HJ, Bakas T, Chumbler NR, et al. A review of the evidence for the use of telemedicine within stroke systems of care: a scientific statement from the American Heart Association/American Stroke Association. *Stroke* 2009;40(7):2616-34.  
[PUBMED](#) | [CROSSREF](#)
17. Cherney LR, van Vuuren S. Telerehabilitation, virtual therapists, and acquired neurologic speech and language disorders. *Semin Speech Lang* 2012;33(3):243-57.  
[PUBMED](#) | [CROSSREF](#)
18. Hill AJ, Theodoros DG, Russell TG, Ward EC. The redesign and re-evaluation of an internet-based telerehabilitation system for the assessment of dysarthria in adults. *Telemed J E Health* 2009;15(9):840-50.  
[PUBMED](#) | [CROSSREF](#)
19. Wang YT, Kent RD, Kent JF, Duffy JR, Thomas JE. Acoustic analysis of voice in dysarthria following stroke. *Clin Linguist Phon* 2009;23(5):335-47.  
[PUBMED](#) | [CROSSREF](#)
20. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE Inst Electr Electron Eng* 1989;77(2):257-86.  
[CROSSREF](#)