Physical Ergonomics & Human Factors, Vol. 223, 2026, 128–133
https://doi.org/10.54941/ahfe1007804

Toward Harmonized Human–Machine Interaction: Assistive Communication for Elderly With Aphasia

Taisuke Sakaki

Kyushu Sangyo University, Fukuoka, 813-8502, Japan

ABSTRACT

This research aims to develop a practical and user-friendly communication device for elderly individuals with higher brain dysfunction, particularly those with aphasia caused by stroke. Current gaze-tracking devices are often ineffective due to unstable eye movements in older users. To address this, the project integrates prior achievements in rehabilitation robotics and elderly monitoring to create functions for facial expression detection, classification, and adaptive learning. Unlike existing emotion-recognition systems, this approach seeks to infer communicative intent from facial features and translate it into speech and synchronized avatar actions, enhanced by adjusted voice frequency, speed, and gestures for clarity. The innovation lies in tailoring the system to individual disabilities, enabling urgent responses (e.g., breathing difficulties), and refining accuracy through machine learning. With Japan's aging workforce—26% of 35 million seniors employed—communication support is increasingly vital to prevent workplace accidents and service decline. This study positions machines not as replacements but as partners harmonizing with human cognition and emotion. By combining assistive robotics expertise with monitoring avatars, the project aims for low-cost, socially implementable solutions applicable at home, in communities, and workplaces, ultimately contributing to safer and more inclusive elderly communication.

Keywords: Assistive communication technology, Machine lip-reading, Elderly care and aphasia, Human-machine interaction, Facial expression-based speech synthesis

INTRODUCTION

Current Status and Challenges of Individuals With Speech Impairments

In Japan, approximately 500,000 individuals experience difficulties in speech, writing, or manual operations due to conditions such as stroke, muscular dystrophy, spinal cord injuries, or severe cerebral palsy (Ministry of Health, Labour and Welfare in Japan, 2026). Despite retaining full cognitive function and awareness, these individuals often face significant psychological distress stemming from their inability to communicate their intentions effectively. This issue is particularly pronounced among older adults, whose unstable gaze control limits the use of commercially available eye-tracking communication devices.

Received March 2, 2026; Revised April 9, 2026; Accepted April 29, 2026; Available online July 20, 2026

While recent advances in AI-based image analysis have enabled the estimation of emotional states from facial expressions, these technologies remain insufficient for decoding actual speech content. Consequently, there is a pressing need to develop new communication devices that allow users to convey their intentions through non-verbal input modalities.

Current State of Machine Lip-Reading Technologies

Research on machine lip-reading has primarily progressed in English-speaking regions, where optical flow and temporal variations in lip shape are commonly used as features to identify phonemes and words based on continuous lip movements during speech. However, direct application of these methods to Japanese is challenging due to fundamental differences in phonological structure. Japanese comprises five vowels and numerous consonant combinations, forming a syllabary of 50 distinct sounds (Midoh, Miura., Inohara, 2024).

For Japanese, approaches have been proposed that estimate vowels from lip images and classify phonemes based on vowel-consonant combinations. These methods often rely on clearly defined lip shapes and involve complex system architectures and computational loads. Some studies have introduced models that define basic lip shapes corresponding to vowels (e.g., /a/, /i/, /u/, /e/, /o/, and closed lips) and reconstruct lip shape sequences from spoken phrases. However, these approaches still face challenges in lip shape detection and sequence processing (Asami, Ishikawa, 2019, Takashima, 2020, Saitoh, 2018).

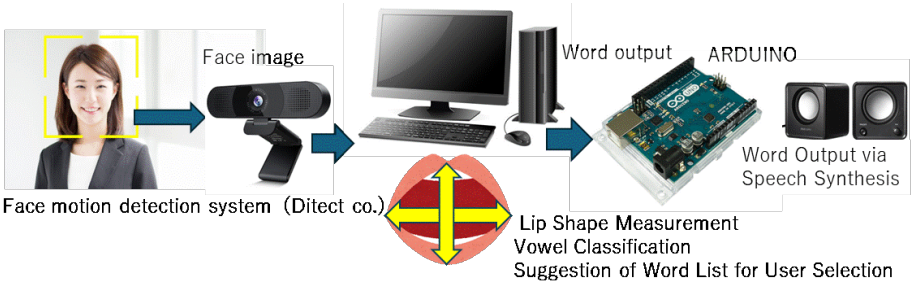

Figure 1: Practical communication support device for elderly individuals with higher brain dysfunction on machine lip-reading techniques for Japanese.

OBJECTIVES AND SYSTEM CONCEPT

Research Objective

The objective of this study is to develop a practical communication support device for elderly individuals with higher brain dysfunction. Specifically, we aim to construct a simplified machine lip-reading system capable of identifying Japanese vowels, thereby contributing to the establishment of communication support technologies for healthy elderly populations as well.

System Overview

The proposed system consists of the following components:

- A front-facing camera connected to a PC for capturing facial images Image Processing: Quantification of lip dimensions (vertical and horizontal), eye angles (vertical/horizontal), and overall facial orientation
- A motion detection system (Ditect Inc., 2026) for real-time measurement of vertical and horizontal lip dimensions Vowel Estimation: Identification of vowels based on lip shape dimensions and facial movements
- A vowel estimation algorithm for identifying six vowel-related mouth shapes (/a/, /i/, /u/, /e/, /o/, and closed lips)
- Text output and synthesized speech generation via Arduino Word Selection: Generation of vowel sequences and presentation of candidate words for user selection
- Speech Output: Vocalization of selected words using speech synthesis

The system includes an automatic threshold calibration function that adjusts to the user's vowel articulation during initialization, based on mean values and standard deviations.

EXPERIMENTAL METHOD

Experimental Design

To evaluate the effectiveness of the proposed system, an experiment was conducted under the following conditions:

- Participants: Two healthy young adult males
- Procedure: Participants were instructed to maintain each of the five Japanese vowels (/a/, /i/, /u/, /e/, /o/) for two seconds, presented in random order, repeated 20 times per vowel
- Measurement: Vertical and horizontal lip dimensions were recorded during articulation, excluding unstable values at the onset and offset of speech
- Analysis: Assuming a normal distribution, the mean and standard deviation for each vowel were calculated

Discriminant Analysis

To assess the distinguishability of vowels, the following method was employed:

- For each vowel, an ellipse was defined using the mean values as the center and the standard deviations of vertical and horizontal dimensions as the axes
- The Euclidean distance between vowel centers was compared with the sum of distances from the center-to-center line to the intersection points of the ellipses

- If the sum of intersection distances exceeded the center-to-center distance, the vowels were considered distinguishable

RESULTS AND DISCUSSION

Summary of Results

The experimental results revealed clear differences in vertical and horizontal lip dimensions across the five vowels. The Euclidean distance between vowel centers was compared with the sum of distances from the center-to-center line to the intersection points of the ellipses. If the sum of intersection distances exceeded the center-to-center distance, the vowels were considered distinguishable. In both trials, the distributions of each vowel were distinctly separated, and their relative positional relationships were preserved. These findings suggest that Japanese vowels can be reliably identified using simple dimensional parameters of lip shape.

CHALLENGES AND FUTURE IMPROVEMENTS

The following issues were identified as areas for future improvement:

- Addressing instability in lip shapes among target users such as stroke patients
- Incorporating auxiliary estimation methods using facial or other body movements
- Expanding the participant pool to evaluate individual variability
- Extending the system to include consonants and continuous speech
- Enhancing robustness against environmental changes (e.g., lighting, posture)

For users with limited ability to change lip shapes, alternative estimation methods based on head orientation or finger movements are under consideration. Additionally, leveraging accumulated data for machine learning could further improve estimation accuracy.

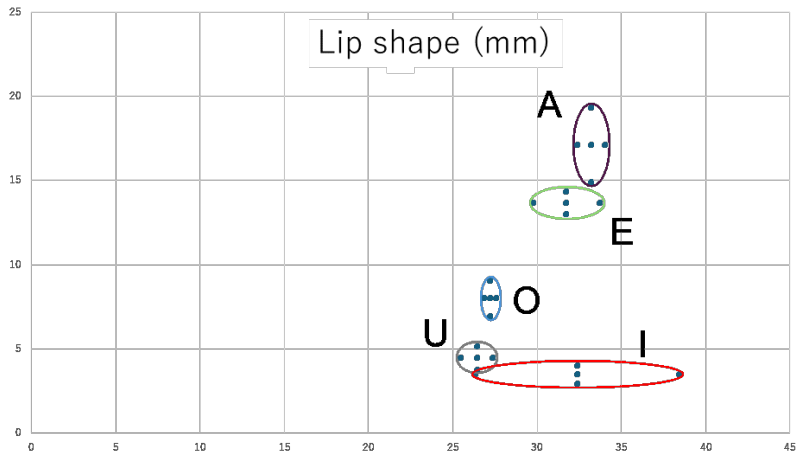

Figure 2: An example of vertical and horizontal lip dimensions for Japanese vowels (/a/, /i/, /u/, /e/, /o/).

Table 1: An example of mean values and standard deviations of horizontal and vertical lip dimensions.

Vowels	Horizontal Mean Value	Horizontal Standard Deviation	Vertical Mean Value	Vertical Standard Deviation
a	33.2	0.82	17.11	2.2
i	32.37	6.03	3.51	0.54
u	26.41	0.94	4.48	0.7
e	31.73	1.97	13.67	0.68
o	27.21	0.35	8.02	1.06

FUTURE PERSPECTIVES

Societal Context and Needs

As Japan faces a declining birthrate and aging population, the employment of older adults is increasing, with 9.22 million (26%) of the 35 million elderly population currently in the workforce. However, this trend is accompanied by a rise in occupational accidents and service quality issues, often attributed to communication barriers. Technologies that support communication among elderly individuals, especially those tailored to varying degrees of impairment, remain underdeveloped and are urgently needed.

Toward Social Implementation

Based on the findings of this study, we aim to develop a low-cost, functionally focused version of the system for practical use. By integrating facial expression analysis for phrase generation, synchronizing with avatar gestures, and adjusting speech frequency and speed, the system can facilitate intuitive communication. Such applications are envisioned for home-based elderly care and workplace interactions among older adults, contributing to the prevention of accidents and service degradation.

Toward Harmonized Human–Machine Interaction and Redefining Human-Machine Relationships: Assistive Communication for Elderly With Aphasia

Traditionally, machines have operated autonomously, requiring humans to adapt to their functions. Moving forward, a paradigm shift is needed toward machines that adapt to human cognition and emotion. This study contributes to this vision by proposing a system that fosters harmonious coexistence between humans and machines through supportive and empathetic interaction design.

ACKNOWLEDGMENT

The authors would like to acknowledge their research staff and the research fund of Kyushu Sangyo University.

REFERENCES

Asami, R., & Ishikawa, T. (2019). "Basic Study on Lip Reading for Japanese Speaker by Machine Learning", IEICE General Conference, Waseda University. https://www.jstage.jst.go.jp/article/jsmermd/2025/0/2025_1P1-F12/_article/-char/en

Ditect inc. (2026) https://www.ditect.co.jp/

Midoh, Y., Miura, N., & Inohara, H. (2024). "Lip2ja: Lip-Reading-Based Japanese Speech System", Osaka University. https://resou.osaka-u.ac.jp/ja/research/2024/20241015_3

Ministry of Health, Labour and Welfare in Japan https://www.mhlw.go.jp/english/

Saitoh, T. et al. (2018). "Research on Multi-Modal Silent Speech Recognition Technology", Kyushu Institute of Technology. https://www.iot.kyutech.ac.jp/wp-content/uploads/2018/07/Impact-Volume-2018-Number-3-June-2018-pp.-47-493.pdf

Takashima, Y. (2020). "Assistive Technology Using Machine Learning Based on Multi-Domain Data for Articulation Disorders", Doctoral Thesis, Graduate School of System Informatics, Kobe University. https://hdl.handle.net/20.500.14094/D1007781