ESTIMATING AMOUNT OF INFORMATION CONVEYED THROUGH SPEECHREADING

HIKI, Shizuo, IMAIZUMI, Kazuya, and FUKUDA, Yumiko*
Graduate School of Human Sciences, Waseda University, Tokorozawa, Japan
**Research Institute, National Rehabilitation Center for the Disabled, Tokorozawa, Japan*

ABSTRACT

A strategy for estimating amount of information conveyed through speechreading and their difference among languages is proposed. In this strategy, first, it is assumed that a set of basic vocabulary involves the same amount of linguistic information in any language. In the process of composing the set of basic vocabulary in each language, trading occurs among selections of vowels and consonants, and compositions of syllables or words. These tradings cause difference among each other language in amount of speechreading information. In order to estimate the amount of information in each stage quantitatively, amount of decrease of information due to the decrease of number of phonetic symbols is calculated. By summing the amounts of decrease of information in a language, it becomes possible to estimate the difference of speechreading information among languages. Charts which relate production and discrimination of vowel and consonant visemes are devised for those estimation.

1. STRATEGY

A strategy for estimating amount of information conveyed through speechreading (lipreading) and their difference among languages is proposed. In this strategy, first, it is assumed that a set of basic vocabulary (consists of 2,000 words, for example) involves the same amount of linguistic information in any language. In the process of composing the set of basic vocabulary in each language, trading occurs among selections of vowels and consonants, or compositions of syllables and words. A language having fewer kind of vowel may use more kind of consonant to compose a syllable, or one having fewer kind of syllable may use more kind of syllable to compose a word. These tradings cause difference among each other language in amount of speechreading information, because effect of speechreading in each stage of trading depends on the phonetic and syllabic system of the language. In order to estimate the amount of information in each stage quantitatively, amount of decrease of information due to the decrease of number of phonetic symbols, from vowels to vowel visemes (mouth shape symbols) and consonants to consonant visemes, is calculated. They are expressed in decrease of entropy (in bits) from the linguistic information of the set of basic vocabulary. In the stage of composition of word, decrease of information due to homophonous word is also calculated. By summing the amounts of decrease of information in all stages of the tradings in a language, it becomes possible to estimate the difference of speechreading information among languages, as the summed decrease is based on the same amount of linguistic information of the set of basic vocabulary. To make the estimation accurate, it is necessary to introduce occurrence and transition probabilities of each linguistic units in each language. This statistical analysis can be done by utilizing an electronic dictionary of each language. In case of Japanese, negative effects of extraordinary large number of homophones is also taken into account [1].

2. MOUTH MOVEMENT

The mouth shapes referred in this paper are three-dimensional contours of upper and lower sides of lips (upper lip / lower lip). They are determined, under constraints by anatomical structure and neural control, by the actions of jaw for open and close and actions of lips for changing in shapes.

In the open and close of the jaw, the angle is changed according to the degree of contraction in muscle groups for open or close.

The contraction of orbicularis oris muscle which draws upper / lower lips forward, downward / upward and inward is main actions in changing the shapes of the lips. In addition to this, there are seven kinds of muscles in upper lip, five kinds of muscles in lower lips. They are divided by their function into the muscle groups which draw upper / lower lips upward / downward, outward and forward. Some muscles among them work for the corners of the lips especially [2].

Originally, the roles of lips are open to take foods in, protrude to suck liquid and close while masticating. But it is considered that finer change of the shapes can be controlled when the lips are applied for utterance of speech. It is thought that pulling the corners of the lips lateralward is supposed to be another roles than eating and drinking [3].

3. DESCRIPTION OF THE SHAPES

A projection of the three-dimensional lip contours onto the frontal view can be approximated to quadratic curve [4]. Consequently, the opening area of the lips is proportional to the multiple of the height in upward and downward direction and the width in lateralward.

The acoustic properties of the vowels are determined by the effective acoustic tube where cross-sectional area and length correspond to the lip opening and depth on mid-sagittal plane. Effective acoustic tube becomes longer if the lips are protruded, while shorter if the corners of the lips are drawn backward. For this purpose, the position of the lips on mid-sagittal plane and corners of the lips must be described as the projections to the lateral view.

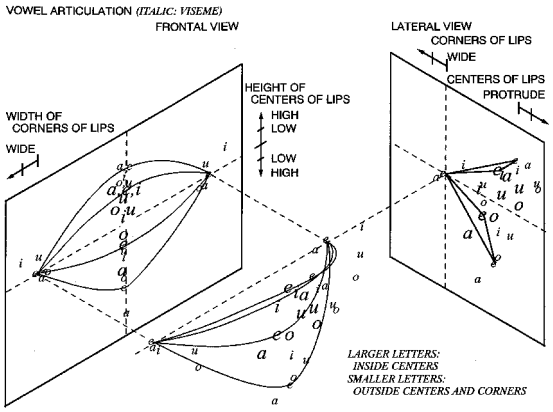

Figure 1. Visually discriminable parts of the space consisted of the scales describing the three-dimensional lip shape on the framework of vowel articulation.

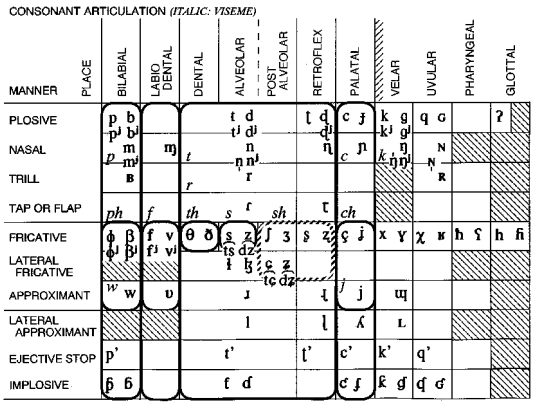

Figure 2. Visually discriminable changes in the mouth shape, categorized on the framework of consonant articulation (IPA).

Thickness of the lips is a part of visual information the mouth shapes. Inside and outside contours of each of the upper and lower lips are necessary to be described.

If the constraints of the anatomical structure and the neural control are utilized, all the three-dimensional shape of the lips may be able to be described uniquely, when the upward and downward height, the lateralward width of the opening of the lips, and depth of the corners are given. The thickness of the lips can be described in the same way.

4. MOUTH SHAPE OF VOWELS

Number of vowels is quite different according to the language, from the fewest three to more than ten kinds in many European and American languages. However, basic three categories [i, a, u] underlay most of the languages.

In Japanese, there are five kinds of vowels [i, e, a, o, u]. A part of basic three vowels [i, a, u] are subdivided, and this is considered to be a standard vowel system in this study. These are discriminable among each other, if the conditions of speechreading are reasonable [5]. Consequently, there are also five kinds of visems corresponding to them.

In the system which supplement insufficient information in speachreading by manual codes in training of hearing speech for hearing impaired children, it is supposed that only three steps of opening; wide, medium and small can be discriminated [6]. In order to discriminate each other vowels, protrusion of the lips and widening of the corners lateralward can be combined, in addition to the three steps of the opening.

5. MOUTH SHAPE OF CONSONANTS

Change of the mouth shapes accompanied by utterance of consonants can be definitely discriminated into group of bilabial plosives [p, b] and bilabial nasal [m], group of labio-dentals (lower lip and upper teeth) [f, v], and group of dental-alveo fricatives (tip of tongue between the upper and lower teeth)[T, D][7]. Place of articulation of those consonants is in frontal part, common to Japanese and European and American languages as English [8].

Furthermore, if the conditions of speechreading are reasonable, group of dental-alveo plosive (upper teeth and tip of tongue)[t, d] and dental-alveo nasal [n], group of alveolar fricatives [s, z], and Japanese flapped [l] can be discriminated in some cases. But, this is not definite because the lips shapes are modified by complicated concatenation rules depending on preceding and following vowels.

In case that the upper and lower teeth or the tip and rear side of tongue can be seen, they are utilized as the discriminating the mouth shapes.

Japanese semivowels [w, y] can also be discriminated, as they are accompanied with the characteristic change of speed to following vowel as clue. For various consonants, too, transient movement peculiar to each of articulation is also useful for discrimination with mouth shapes.

REFERENCES

[1] Hiki, Shizuo, and Kamikubo, Emiko. Comparison of effect of homophones on speechreading between Japanese and English, *Journal of the Acoustical Society of America,* Vol. 100, No. 4, Part 2, pp. 2790-2791, 1996 (Program of the 3rd Joint Meeting of the Acoustical Society of America and Japan).

[2] Hiki, Shizuo, and Harshman, Richard. Speech synthesis by rules with physiological parameters, *Journal of the Acoustical Society of America,* Vol. 46, Supplement 1: p. 111, 1969 (Program of the 77th Meeting of the Acoustical Society of America).

[3] Ohala, John J. The acoustic origin of the smile, *Journal of the Acoustical Society of America,* Vol. 68, Supplement 1: p. 33, 1980 (Program of the 100th Meeting of the Acoustical Society of America).

[4] Fromkin, V. Lip positions in American English vowels, *Language and speech*, Vol. 7, Part 4, Oct./Dec., pp. 215-225, 1964.

[5] Fukuda, Yumiko and Hiki, Shizuo. Characteristics of the mouth shape in production of Japanese: Stroboscopic observation. *Journal of the Acoustical Society of Japan (B)*, Vol. 3 No. 2: pp. 75-91. 1982.

[6] Henegar, Mary Elsie and Cornett, R. Orin. Cued Speech Handbook for Parents. Cued speech program, Gallaudet College. 1971.

[7] Hiki, Shizuo. Possibilities of compensating for defects in speech perception and production. *Proceedings, 1994 International Conference on Spoken Language Processing (ICSLP'94),* September 18-22, 1994, Yokohama, Japan, Vol. 4, pp. 2245-2252 (Plenary Lecture).

[8] Kircos, P. B. and Sharon, L. S. Differences in visual intelligibility across talkers. *The Volta Review*, Vol. 84, No. 219, 1982.