ARTICULATION CHANGES IN DIFFERENT VOICING PATTERNS.

Kiyoshi Honda, Shinji Maeda, and Miyoko Sugito

Univ. Paris III (France) and ATR-CIS (Japan), CNRS and ENST (France), and Institute for Speech Communication Research (Japan)

honda@atr.jp, maeda@tsi.enst.fr, and msugito@nifty.com

ABSTRACT

Human speech arises from orchestrated activities of phonatory and articulatory organs and reflects human-specific characteristics in anatomy and physiology. The tongue and larynx are less tightly coupled in humans, and they are also innervated separately from the cortex. These biological specificities provide aerodynamic and acoustic bases of speech production and contribute to generating a parallel time-pattern of gradually changing vocal signals with ripples in amplitude and spectrum due to rapid articulatory movements. A close look at local sound variations suggests that tongue-larynx linkage still exists as an old trait common to the primate family, as seen in the variation of vocal frequency due to articulation. Contrarily, articulatory control may also be influenced by laryngeal control, as seen in irregular articulation in certain vocal expressions. Vowel devoicing may be a complex case of such bilateral interactions, and a special attention was made on the topic in this report.

Keywords: Articulation and voicing, intrinsic vowel features, vowel devoicing.

1. INTRODUCTION

Vowels exhibit dual acoustic patterns in speech: formant patterns that derive from articulatory settings, and intrinsic prosodic features that arise from the influence of articulation. The intrinsic features are vowel-specific variations in intensity, duration, and fundamental frequency (F0) [1], and they together complete the nature of vowels in speech. These features indicate biological evidence that the forms and functions of speech production system interact with each other in the multiple domains.

Recent MRI data recorded for word utterances with voiced and devoiced vowels exhibited a small difference in tongue shapes for vowel /i/. This suggests that production of the voicing contrast can affect vowel articulation through a certain process. How voicing contrast influences supralaryngeal articulation is an interesting question, but it is of course not readily explained by the biomechanical account for the effect of articulation on voicing, i.e., intrinsic vowel F0.

This report describes production processes of the intrinsic features of vowels by taking into account the interaction processes among speech production subsystems and discusses possible mechanisms of changes of articulation in different voicing types using a case of vowel devoicing as an example.

2. BASIC ISSUES

A few anatomical and physiological issues will be reviewed here before discussing the topic of articulatory changes due to voicing conditions.

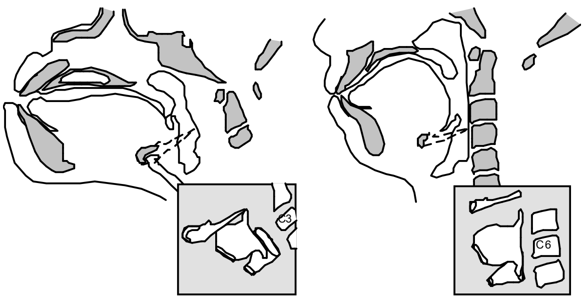
2.1. Human-specific System

Human speech relies on several anatomical and physiological characteristics of the speech organs. Two major characteristics are the functional separation of the larynx from other speech organs [2] and the direct neural pathway from the cortical motor area to the motor nuclei in the brainstem [3]. They both contribute to the sound patterns of sequence of syllables on vocal melody and sequence of words in a single breath. An additional factor that derives syllable-specific articulatory patterns is the fact that the speech motor system is composed of slow and fast organs. The slower articulators produce the core of a syllable, and the faster organs can modify the syllable before and after the core.

2.2. Comparative Anatomy

Comparative anatomy of the vocal organs of the human and nonhuman primate points to human-specific forms: reduced oral cavity length, elongated pharynx, round tongue, low larynx position, and thus the separation of the tongue and larynx. Figure 1 shows midsagittal MRI tracings of a male chimpanzee and a male human. In non-human primates, the arrangement of the vocal organs indicates a tight coupling of the tongue and larynx because the hyoid bone and thyroid cartilage are mechanically unified. In humans, on the other hand, the thyroid cartilage separates from the hyoid bone, and this change enables independent control of tongue movements and laryngeal functions. The separation of the larynx from the articulators comprises the anatomical background of the source-filter function in speech production processes [4], but this separation is incomplete, and soft-tissue connections among the organs via the hyoid bone limit functional separation between articulation and voicing [5].

Figure 1: MRI tracings of a male chimpanzee and a human male, showing differences in the articulators' geometry and laryngeal framework.



(a) Chimpanzee (b) Human

2.3. Mandible, Tongue, and Larynx

The uniqueness of human jaw action is the translation of the temporomandibular joint (TMJ). In non-human primates, the jaw opens wide only by hinge movements because there are no rigid obstacles behind the jaw. On the other hand, the human jaw opens not only by rotation but also by forward translation. The position of the human TMJ is as high as the level of the external ear canal, and the well-developed tympanic bone and narrow retromandibular fossa inhibit a large jaw opening by pure rotation. When the jaw opens, the mandible comes closer to the tympanic bone. Then, the center of jaw rotation shifts from the TMJ to the contact point of the mandible and tympanic bone. The mandible offers the rigid framework to the tongue and larynx to interact with vowel articulation. In high vowels, the jaw opens wider for higher F0 with further forward translation [6]. This translation compensates for F0 lowering due to the jaw opening by the same tongue-larynx connection that causes their high intrinsic F0.

2.4. Acoustic and Aerodynamic Factors

There are a few acoustic and aerodynamic effects that are less noticed to exist but should be considered in vowel production. Possible factors causing the differences between natural and synthetic vowels are the transvelar nasal coupling [7] and subglottal air pressure variation across vowels [8]. High vowels have the greater effect of transvelar coupling, which increases nasal output because of the higher intraoral air pressure. Subglottal pressure appears to show these variations in a way to compensate for the low intensity for high vowels. These effects can be predicted by evidence that synthetic high vowels from a vocal-tract model show extremely small amplitude.

3. INTRINSIC FEATURES OF VOWELS

It has been known that vowels exhibit three intrinsic tendencies in intensity, duration, and F0. Each feature appears to occur at a certain specific condition: e.g., intrinsic intensity can be seen in isolated productions, while intrinsic F0 is more evident in words. In what follows, these intrinsic features are explained in relation to the utterance forms (V, CV, CVC, and CVCV). Vowel devoicing in Japanese is discussed here because the same mechanisms appear to underlie the devoicing and intrinsic features.

3.1. Intrinsic Intensity in V

High vowels have lower intensity, which may be observed in all utterance types including isolated vowel production. This variation is due to acoustic and aerodynamic consequences in the vocal tract. A strong vocal-tract constriction limits acoustic output, and elevation of intraoral pressure reduces transglottal pressure difference. In high vowels, the transvelar nasal coupling and subglottal pressure elevation may compensate for the small amplitude of high vowel sounds, as described before.

3.2. Intrinsic Duration in CV

High vowels have shorter duration than low vowels. In high vowels, the buildup of vocal-fold vibration becomes slower due to the reduced transglottal pressure difference, and the vowel duration becomes shorter. This may also be augmented by the higher vocal-fold stiffness for high vowels, as described in the next section.

3.3. Intrinsic F0 in CVC

High vowels tend to have higher F0 than low vowels when they are produced in words. Many accounts have been proposed for this tendency, and the following is one of the biomechanical accounts. The entire larynx is suspended from the hyoid bone by muscles and ligaments. When the hyoid bone is advanced by contraction of the tongue muscles used for high vowel articulation, the thyroid cartilage rotates forward and the vocal folds are stretched, causing a F0 rise. The larynx is also connected to the mandible thought muscles and ligaments partly via the hyoid bone. Jaw opening movement for low vowels compresses the space between the mandibular symphysis and the cervical spine, backing the hyoid bone and rotating the thyroid cartilage to lower F0. Therefore, there are two causal factors, one for high vowels to have higher F0, and another for low vowels to have lower F0.

3.4. Vowel Devoicing in CVCV

Vowel devoicing often occurs when high vowels are surrounded by voiceless consonants. The devoicing involves laryngeal reorganization of glottal abduction control. Its causal mechanism may be supplemented by the following three factors that also underlie beneath the vowel intrinsic features described above.

- Acoustic factor – vocal-tract constriction for high vowels reduces syllable sonority and characterizes the amplitude profiles of the syllable with high vowels in words or sentences.
- Aerodynamic factor – vocal-tract constriction for high vowels and vocal-tract wall stiffening for voiceless consonants together increase the intraoral pressure and reduce the transglottal pressure difference, which gives rise to the condition that inhibits vocal-fold vibration.
- Biomechanical factor – higher vocal-fold tension for high vowels is an additional condition that inhibits vocal-fold vibration. Tongue-larynx interaction for intrinsic F0 and cricothyroid muscle activity for voiceless consonants together contribute to increasing vocal-fold tension.

4. DEVOICED VOWEL IN JAPANESE

In Japanese, the frequency of vowel devoicing is known to depend on dialects. In Tokyo dialect, high vowels in voiceless consonant environment are regularly devoiced, while in Osaka dialect, they occur only irregularly [9]. This dialectal difference is often explained by phonologization process. In Tokyo dialect vowel devoicing is phonologized to be observed consistently, while in Osaka dialect it is not. It is interesting to observe experimentally the differences of articulatory positions between the cases with and without vowel devoicing. The authors conducted a motion MRI experiment to record the movement of the speech organs in the two conditions.

4.1. MRI on Devoiced and Undevoiced Vowels

Motion MRI experiments were conducted using a synchronized sampling imaging with external triggering [10]. Subjects listen to the trigger sounds of a repetitive rhythm to repeat word utterances with and without devoicing high vowels in consonantal environments. Data were recorded from two male speakers of Tokyo dialect and Osaka dialect. The utterances used are /kiki/ and /kika/ with a lexical accent on the first syllable that is subjected to devoicing.

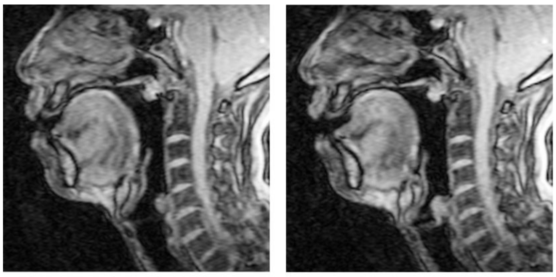
4.2. Articulatory Changes in Devoicing

Figure 2 shows the frames for the first vowel in the word /kika/ from the Tokyo dialect speaker with devoiced and undevoiced conditions. When the vowel /i/ is not devoiced, the tongue shape takes a complete form of the vowel, tight palatal constriction with an oval shape of the tongue body. When the vowel is devoiced, the tongue body is more rounded, and the mandible is elevated. Also in devoiced cases, tongue movement during the first syllable /ki/ is limited, and it appears as if /k/ is prolonged. Thus, the articulatory settings for the devoiced vowel /i/ can be explained with respect to assimilation to the initial consonant.

4.3. What are the Causal Factors?

Devoiced vowels in Japanese have been studied with laryngeal fiberscopy, and a single abduction movement of the glottis has been the common observation for the CVC segments [11]. Articulatory contrasts in devoicing have not been examined until very recently, and the result shown in Fig. 2 is rather surprising because the devoiced vowel is realized by a consonant-like glottal gesture and a tongue shape similar to the initial consonant. However, this observation alone does not explain the random occurrence of vowel devoicing in Osaka dialect. Based on the underlying factors for vowel devoicing described above, it may be speculated as follows. The glottis opens wide through the initial consonant and vowel, and therefore the high intraoral pressure tends to be maintained during the constricted vowel. Supposing that the intraoral pressure comprises a sensory target for the vowel, tongue articulation may adjust to the random variation of intraoral pressure drop to reduce tight constriction in the anterior vocal tract.

Figure 2: MRI data from a speaker of Tokyo dialect during utterances /kika/ with a high-to-low accent type.



5. DISCUSSION

Biomechanical or aerodynamic processes in realizing some intrinsic vowel features involve dynamic adjustments among structures, which causes specific changes while maintaining the basic vowel features. These adjustments are partly supported by the redundancy of the motor system with multiple means to achieve the set of targets. In producing high vowels, the tongue and jaw work together to adjust tongue height. In /i/ with high F0, the excessive force from the contraction of the genioglossus muscle can be compensated for by jaw lowering. In low vowels with high F0, jaw opening with marked forward translation realizes high F0 with wide jaw opening. A recent MRI study showed that the transverse muscle cooperates with the geniolgossus to rapidly raise the front tongue [12], which suggests a redundant control mechanism within the tongue.

In CVCV utterances with non-aspirated voiceless consonants, the two successive syllables tend to show different degrees of glottal opening. In Japanese utterances, at least, the first consonant shows the wider glottal opening than the second one, even when the word is produced with a flat accent pattern [11]. This phenomenon has been known for long, while plausible accounts are not available to date. Considering the tendency of vowel devoicing in such a syllable sequence, it may be that the glottis abducts wider to allow a higher airflow for a word-initial consonant followed by gradual glottal constriction. Or, it may be that a certain sensory feedback mechanism is involved in the utterance control: the motor system automatically adjusts to realizing the intraoral or subglottal pressure pattern. This interesting question must be examined by articulatory and aerodynamic measurements in future experiments.

Acknowledgement

This work was partly supported by the Ministry of Education, Science, Sport and Culture, Grant-in-Aid for Scientific Research (A), 16202006, 2004–2007.

REFERENCES

[1] Peterson, G.E., Barney, H.L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am.* 24, 175-184.
[2] Fink B.R. 1975. *Human Larynx: A Functional Study*. New York: Raven Press.
[3] Jürgens, U. 1979. Neural control of vocalization in non-human primates. In Steklis, H.D. & Taleight, M.J., (eds), *Neurobiology of Social Communication in Primates: An Evolutionary Perspective*, Academic Press, 11-14.
[4] Fant G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.
[5] Honda, K. 1995. Laryngeal and extra-laryngeal mechanisms of F0 control. In Bell-Berti, F. & Raphael, L.J. (eds), *Producing Speech: Contemporary Issues*, New York: American Institute of Physics, 215-232.
[6] Erickson, D., Fujimura, O. 1996. Maximum jaw displacement in contrastive emphasis. *Proc. 4th Int. Conf. Spoken Language Processing (ICSLP 96)*, 141-144.
[7] Dang, J., Honda, K. 1996. An improved vocal tract model of vowel production implementing piriform resonance and transvelar coupling. *Proc. 4th Int. Conf. Spoken Language Processing (ICSLP 96)*, 965-968.
[8] Bucella, F., Hassid, S., Beeckmans, R., Soquet, A., Demolin, D. 2000. Pression sous-glottique et débit d'air buccal des voyelles en français. *Actes des XXIIIèmes Journées d'Etudes sur la Parole*, Aussois, pp. 449-452.
[9] Sugito, M. 2003. Timing relationship between prosodic and segmental control in Osaka Japanese word accent. *Phonetica*, 60, 1-16.
[10] Masaki, M., Tiede, M. K., Honda, K., Shimada, Y., Fujimoto, I., Nakamura, Y., Ninomiya, N. 1999. MRI-based speech production study using a synchronized sampling method. *J. Acoust. Soc. Jpn. (E)*, 20, 377-381.
[11] Sawashima, M., Hirose, H. 1981. Abduction-adduction of the glottis in speech and voice production. In Stevens., K. N., & Hirano, M. (eds), *Vocal Fold Physiology*, Tokyo: University of Tokyo Press, 329-346.
[12] Takano, S., Honda, K. 2006. "Measurement of tissue deformation in the tongue during a vowel sequence /ei/ using tagged-cine MRI," In Proceedings of the 7th International Seminar on Speech Production, Ubatuba-SP, Brazil, pp 63-71.