

Perceptual cues to vowel quantity: Evidence from Swedish and Japanese

Takayuki Arai*, Dawn Behne†, Peter Czigler‡, and Kirk Sullivan‡

*Sophia University, Tokyo, Japan

†Norwegian University of Science and Technology, Trondheim, Norway

‡Umeå University, Umeå, Sweden

Abstract

Research on the perception of Swedish vowel quantity suggests that when the duration of a vowel is relatively long, due for example to inherent duration or postvocalic voicing, vowel quantity might not be adequately cued by duration alone and might also make use of the vowel spectra to distinguish vowel quantities. The current project investigates whether Japanese listeners use spectral cues to identify the quantity of vowels which have a relatively long inherent duration. Results are compared with Swedish findings.

Introduction

Background

The vowel systems of some languages are described as having contrastive vowel quantities. Vowel quantity refers to the phonological distinction of a vowel relative to one or more other vowels of similar timbre in the language. Contrasts in vowel quantity are often acoustically realized by the duration of vowels, with a long vowel quantity having a duration which extends over more time than a short vowel quantity. The greater amount of time associated with a long vowel quantity also allows the possibility for a more extreme articulation than a corresponding short vowel quantity. Consequently, the vowel spectrum, in particular the first and second formant frequencies, and perceived timbre (e.g., Stevens & House, 1955) may also be affected by vowel quantity.

Swedish has traditionally been characterized as having distinctions between long and short vowel quantities (Elert, 1964). In a classic study on Swedish, Hadding-Koch and Abramson (1964) concluded that vowel duration was a primary perceptual cue to Swedish vowel quantity, but could not exclude the possible influence of the vowel spectra.

Recent studies (Behne, Czigler & Sullivan 1996, 1998) have reexamined the effects of vowel duration and the first two formant frequencies on perceived vowel quantity identification in Swedish. Based on natural productions, materials were synthesized with 10 stepwise adjustments of vowel duration and 10 stepwise adjustments of F1 and F2 (a total of 100 synthesized words) for each of three vowel quantity pairs. This was done for /kVt/ (Behne et al, 1996) and /kVd/ (Behne et al, 1998) words. The results for /kVt/ words illustrated that Swedish listeners use vowel duration, but not spectral information, to identify the quantity of the vowel pairs [i:]-[ɪ] and [o:]-[ɔ] whereas for the vowel pair [A:]-[a], listeners use both vowel duration and spectral attributes of the vowel. In the case of the /kVd/ items, listeners were even more likely to use both vowel duration and spectra to identify vowel quantity.

The production of the vowels [A:] and [a] involves moving the relatively heavy jaw, resulting in a longer inherent vowel duration than the other vowels studied (e.g., Peterson & Lehiste, 1960). Postvocalic voicing is also known to cause the duration of a preceding vowel to be longer in Swedish and other languages (e.g., House & Fairbanks, 1953). As a possible explanation for our results, we considered that it may simply be difficult to lengthen

relatively long vowels indefinitely without disturbing the natural rhythm of speech; when the duration of a vowel is relatively long (due, for example, to inherent duration or postvocalic voicing), it may be more likely to reach the extent to which a duration is vaguely permissible for a vowel as a rhythmic unit within a word. As a consequence, vowel quantity may not be adequately cued by vowel duration alone and might make use of the vowel spectra too. If so, listeners would, in those cases, be more dependent on secondary acoustic cues, such as the vowel spectrum.

Current Study

If this motivation for the Swedish findings is well founded, we would expect to find the same general pattern of results in a language such as Japanese, which distinguishes vowel quantities but is also unrelated to Swedish.

As an initial investigation of this hypothesis, the current study was carried out to examine how Japanese listeners use vowel duration and the first two vowel formant frequencies when identifying Japanese vowel quantity, and to compare it with the Swedish listeners' responses from the earlier studies (Behne et al, 1996, 1998). Of particular interest is whether Japanese listeners are more likely to use spectral cues to identify the quantity of the inherently long vowels, /a:/ and /a/, compared to other vowel pairs.

Method

The recordings, measurements, synthesis and experimental procedures used here are closely matched with those in (Behne et al, 1996, 1998). For consistency, the measurements and synthesis (sections 2.12. and 2.13.) were carried out at Umeå University where the Swedish materials had been prepared. The experiment running program used for the identification task (section 2.2) was based on the program originally used for the Swedish study.

Materials

Recordings. A set of six /kV/ real words containing the vowels [i, i:, o:, o, a:, a] were used as targets. Since the vowel in /ki/ is commonly devoiced, the speaker was asked to produce the devoiced variants and also try to produce voiced variants.

A adult phonetically-trained native male speaker of the Tokyo dialect of Japanese was recorded producing 10 random repetitions of the target words in the sentence "Mou ichido ___ to iu tango-wo itte kudasai." ("Please say the word ___ again"). The productions were made at his natural speaking rate and using a natural intonation. Care was taken to avoid a falling pitch with long vowel quantities.

Measurements. From the 10 productions of each target word, ESPS/waves+™ was used to measure the vowel duration and the first three formant frequencies of the vowel, measured at the center of the vowel's most evident steady state. For each repetition of the six target words, the mean value of these measures was calculated and the production which best corresponded to the mean values was chosen to be used as the basis for resynthesis.

Among the productions of /ki/ which were expected to have a devoiced vowel, voicing was observed in two of the 10 cases. The vowel duration and formant frequencies of these 2 items were comparable to the 10 productions of /ki/ with the voiced vowel. Consequently, items used for resynthesis were chosen from among variants of /ki/ with the voiced vowel.

Synthesis. Using the Kay Elemetrics LPC Parameter Manipulation/Synthesis program, the measurements of the selected productions were used as extreme points of a 10x10 synthesis matrix, having ten degrees of vowel duration and ten degrees of simultaneous first and second formant frequency adjustment. Starting from the selected production of [ki:], [ko:] and [ka:], the vowel duration and together, F1 and F2, were adjusted in equal-sized steps toward the measured postvocalic closure duration of the selected productions of [ki], [ko] and [ka] respectively. In each series the third formant frequency of the vowel was kept the same as it had been in the long vowel quantity.

Identification task

Twenty-four native, young, adult, speakers of the Tokyo dialect of Japanese participated in the study. All were affiliated with Sophia University in Tokyo at the time of the study.

Subjects were seated wearing headphones at a computer terminal with a monitor and mouse. For each trial, subjects heard a synthesized word and two /kV/ words were presented on the monitor in katakana. The words on the monitor differed in vowel quantity and had the same vowel quality as the target words which the synthesized items in that series were based on. Subjects were instructed to use the mouse to click on the visually presented word which they heard. They were asked to respond as quickly as possible and were allowed up to 10 seconds to respond before the beginning of the next trial, although subjects rarely encountered this upper limit. Subjects heard a total of 1500 randomized items (3 vowel qualities x 100 items x 5 repetitions). Before starting the experiment, subjects had three practice trials, and after each set of 50 trials, subjects had the occasion to take a short break.

Results and Discussion

Listeners' responses (Figure 1) and reaction times across the 10 vowel and spectral steps demonstrate a clear pattern. For all three word pairs, results show listeners using vowel duration, but not spectral information, to identify a vowel's quantity.

These findings are consistent with general observations of how vowel quantity is acoustically realized across languages and earlier investigations of vowel duration as a perceptual cue to vowel quantity in Japanese (e.g., Fujisaki, Nakamura & Imoto, 1975). Based on results from Swedish (Behne et al, 1996, 1998), it was also expected that vowel duration would offer a primary perceptual cue for the vowel quantity identification in Japanese. We had further hypothesized that listeners would tend to make use of spectral information in cases when the vowel was relatively long due to other linguistic factors. In particular, with the materials used in this study, we were interested in whether listeners would be more likely to use the vowel spectra to identify the quantity of the inherently longer low Japanese vowels, /a:/ and /a/. The results showed no indication of this pattern.

Notably, the materials used in the present study may be considered a strict context for observing the pattern earlier found for Swedish. The closed syllable (/kVt/ and /kVd/) used in the Swedish studies would put greater restrictions on the possible vowel duration of the syllable than the open syllable (/kV/) used in the Japanese study. This may have resulted in a phonetic context in the Swedish materials where vowel quantity could not be adequately cued by vowel duration alone and that corresponding syllable-internal limitations on vowel duration were not present in the Japanese study.

The Japanese results are, however, consistent with previous research, including the findings for Swedish. These findings suggest a complex, but systematic, role of vowel duration and spectra in distinguishing vowel quantity and set us in the direction of a follow-up study to

further examine effects of syllable structure on the use of vowel duration and spectral information as perceptual cues to vowel quantity.

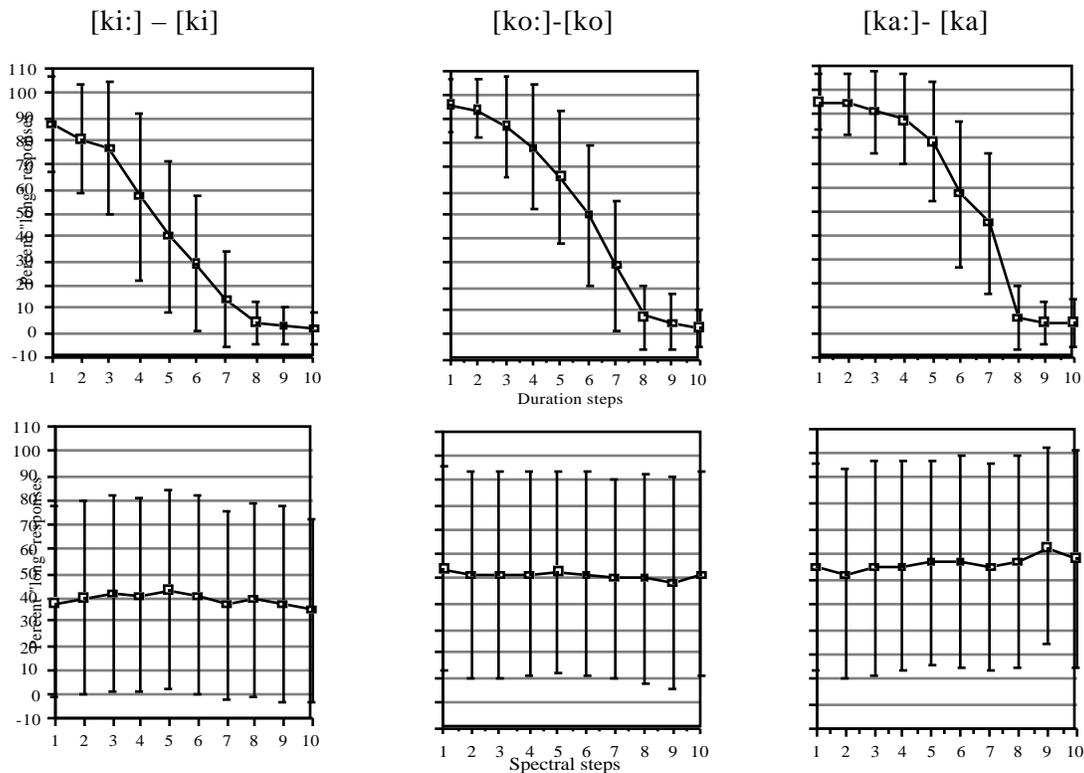


Figure 1. Mean percent long responses are plotted for the 10 synthesized duration steps and spectral steps for the three pairs: [ki:]-[ki], [ko:]-[ko] and [ka:]-[ka].

Acknowledgements

The authors thank Yuko Mimura, Kenji Okada, Ola Andersson, and Thierry Deschamps for their programming and technical assistance, and the listeners for their patience.

References

- Behne D., Czigler P. and Sullivan K. 1996. Acoustic characteristics of perceived quantity and quality in Swedish vowels. *Speech Science and Technology '96*, 6, Adelaide, 49-54.
- Behne, D. M., Czigler, P. E., and Sullivan, K. P. H. 1998. Perceived vowel quantity in Swedish: Effects of postvocalic voicing. *Proceedings of the 16th International Congress of Acoustics and the 135th Meeting of the Acoustical Society of America*; 2963-64, 1998.
- Elert C-C. 1964. *Phonologic studies of quantity in Swedish*. Stockholm: Almqvist & Wiksell.
- Fujisaki, H., Nakamura, K., and Imoto, T. 1975. Auditory perception of duration of speech and non-speech stimuli. In Fant, G. and Tatham, M. (eds.) *Auditory analysis and perception of speech*. London: Academic Press, 197-219.
- Hadding-Koch K. and Abramson A. 1964. Duration versus spectrum in Swedish vowels: some perceptual experiments. *Studia Linguistica* 2, 94-107.
- House A., and Fairbanks G. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *Journal of the Acoustical Society of America* 25, 105-113.
- Peterson G and Lehiste I. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32, 693-703.
- Stevens K. and House A. 1955. Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America* 27, 484-493.