Weighted Constraints in Phoneme Perception

石川, 潔 / ISHIKAWA, Kiyoshi

(出版者 / Publisher)
法政大学文学部
(雑誌名 / Journal or Publication Title)
Bulletin of Faculty of Letters, Hosei University / 法政大学文学部紀要
(巻 / Volume)
48
(開始ページ / Start Page)
75
(終了ページ / End Page)
81
(発行年 / Year)
2003-03-03
(URL)
https://doi.org/10.15002/00003982

Weighted Constraints in Phoneme Perception*

Kiyoshi ISHIKAWA

abstract

The roles played by two allophonic constraints in the perception of stop consonants in Japanese are discussed. If the finding reported in this paper is real, it means that different constraints are utilized with different strengths.

Keywords: phoneme perception

1 Introduction

It is not a new claim that constraints (or principles or rules or whatever) are associated with different weights. For example, in GB syntax, ECP violations were assumed to lead to more severe unacceptability than Subjacency violations. A similar situation obtained in pre-OT phonology too.[1] These proposals are different from OT, in which, while constraints are ranked and hence violable, the output is either totally acceptable or unacceptable; OT is not meant to account for degrees of unacceptability.

The purpose of this paper is to report a non-experimental finding which suggests the need for a somewhat different conception of constraint weight. The discussion concerns the roles played by two allophonic constraints in the perception of stop consonants in Japanese. The finding is that one of them is more readily utilized than the other. If this finding is real, it indicates that constraints have different strengths or weights from each other.

---

* This paper is a by-product of my 2002 undergraduate seminars. I would like to thank the participants.

(1) I have in mind the dicussion of long sequences of stressless syllables in Fromkin (2000: 630), for example.

2 Formant Transition Cues

Consider the following English words, as spoken in isolation:

(1) a. heat
b. heap

which are schematically of the form:

(2) ... VC

The word-final stop consonants are sometimes unreleased, in which case the only phonetic cue available in perceiving the words is the formant transitions of the vowels.[2] Schematically, the "rule" for the formant transition can be stated as something like:

(3) ... VC ⟶ ... f (V,C)C

where f is a function from a vowel and a consonant to a vowel with an appropriate formant transition. Now, is this "rule" merely a natural result of human vocal tract, or a part of the speaker's linguistic knowledge in addition?

The fact that native speakers of English can utilize the transition to identify the physically absent stop consonant suggests that it is part of his/her phonetic knowledge. On the other hand, native speakers of Japanese do not seem to be able to utlize the formant transition to identify the following stop consonant in the environment (2). This initially seems to suggest that it is *not* part of the phonetic knowledge of a native speaker of Japanese. However, Kakehi et al. (1996) argue that such cues *are* utilized by speakers of Japanese at the initial stage of phoneme perception (the phoneme cues extraction stage), which means that something like (3) is part of the phonetic knowledge of Japanese.

---

(2) It was Takehiko Makino who taught this, when I was a complete novice in phonetics (back in 1996).

Their argument is based on their experimental finding of the following kind. They presented subjects with phonetic stimulus of the following form:

V_1CV_2

where V_1 and V_2 are vowels and C is one of the stop consonants /p, t, k/. C physiologically imposes particular formant transitions on V_1 and V_2, which could be utilized as phonetic cues for C. Now, if native speakers of Japanese do not have something like (3) as part of his or her phonetic knowledge, then masking the relevant formant transition of V_2 (together with the release burst) with noise should disable their identification of C. However, their subjects' success to identify C in such a situation was found to be statistically significant, thereby suggesting that they do have something like (3) as part of their knowledge. They attribute the subjects' failure to identify C in (2) to Japanese phonotactics: (3) is a permissible syllable, but (2) is not.[3]

Recently I have attempted to reproduce their results, with no success. The failure can be attributed to the following three reasons: (i) I was trying to obtain their results with novice listeners, while their subjects were "employees of NTT who were trained in taking speech perception tests" (p.127), (ii) I was aiming for a 100% success, while their results were only statistical and with none of their stilumi were their subjects able to reach a 100% correctness, (iii) I was trying to obtain the result with the same masked vowels throughout the stimuli, while they did not. The last point needs to be elaborated a bit further. Consider the following three sequences:

---

(3) I'm not concerned with the exact formulation of their account. For example, if (i), an instance of so-called gemination, should be syllabified as indicated here,

(i) kek.ka

one could argue that (2) should count as a legitimate syllable in Japanese. However, at the phoneme level, (i) is something like:

(ii) keQ.ka

The coda [k] appears as a result of the onset [k] of the following syllable, and when nothing follows, /Q/ surfaces as a glottal stop, not [k]. Thus, [kek] in isolation does violate Japanese phonology. Kakehi et al.'s point is that Japanese speakers' utilization of phonetic cues follows what Japanese phonology dictates, and whether the relevant phonological constraint is phonotictics at the phoneme level or allophonic variation is not the issue.

(4) a. /ape/
b. /ate/
c. /ake/

They produced their stimuli by masking the initial part of /e/ with noise of different lengths. The masked vowels to be heard after /ap/, /at/ and /ak/ in the stimuli were produced from the /e/'s in (4a–c) respectively. In other words, different masked vowels were used in different cases. In contrast, I employed the same single masked vowel (produced, say, from (4a)) in all cases. If the masked /e/ still retained the formant transition caused by the preceding stop consonant, the listener's success in identification of the medial stop consonant could be partially due to the masked /e/, instead of the preceding /a/. In short, I was trying to reproduce their result under more severe conditions and failed.

3 Coda Nasal Allophones

Incidentally, during such an attempt, I happened to have produced an interesting stimulus:

(5) Konpa-de kenka sita.
party-at fight did
'(I) had a fight at the party.'

In Japanese, the coda nasal exhibits the following allophonic variation:

(6) a. Realized as [m] before bilabials.
b. Realized as [ŋ] before velars.

Thus, phonetically, (5) is pronounced as:

(7) kompa-de keŋka ʃita

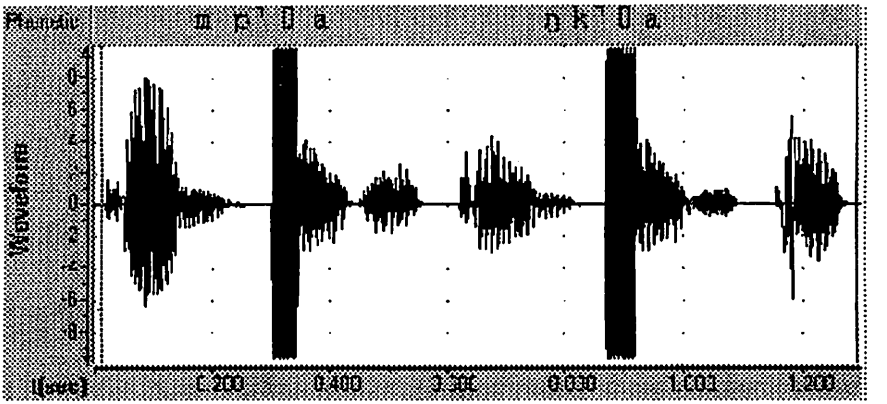

figure 1: The masked version of (7)

I first recorded my own pronunciation of (7) and then masked the initial part of [a] in *konpa* (again, together with the release burst of [p]) with noise (or a beap sound), the result of which is a beap sound followed by a vowel. The [pa] in *konpa* and the [ka] in *kenka* are both replaced with this sequence, the result of which is shown in Figure 1.[4] The beap sound is indicated by □ and the two [a]'s indicated are exactly the same; thus, this stimulus obeys the stricter condition with respect to (iii).

Surprisingly, this stimulus was perceived as (7) not only by my own ears, but also by everybody else with whom I have informally tested it, the undergraduate students in my seminars, and even by those high school pupils who visited our campus when we hosted an advertisement campaign of our department (those who must have received absolutely no training in phonetics or speech perception tests!). Although these are not a result of a controlled experiment, they suggest that this stimulus obeys the stricter condition with respect to (i) and (ii) too. These results suggest that, while both (3) and (6) are real constraints and part of the native speaker's knowledge, the latter are more readily utilized than the former.

---

(4) Speech Analyzer Ver. 1.06a (by Summer Institute of Linguistics) was used for recording and displaying the sound wave. TWE Ver. 2.3.1 (by Yamaha) was used for editing.

(8) Nasal allophonic constraints are more readily utilized than the pre-stop formant transition constraints.

Indeed, this stimulus is a grammatical sentence, instead of non-words, and hence there is the possibility that top-down processing helped the perception.[5] In addition, the results are not obtained from a controlled experiment. Thus, further work is needed to establish (8). However, if real, (8) can be understood as indicating that different constraints are "entrenched" to varying degrees; a more deeply entrenched constraint is more readily utilized.

4 Implications

My speculation is that, because the inputs that speakers of Japanese confront everyday all conform to the Japanese phonotactics, the only opportunities in which they (have to) utilize the pre-stop formant transition constraints, instead of the post-stop formant transition constraints, are those in which the post-stop cues are unavailable for one reason or another (say, when they non-artificially ended up masked). Thus, the pre-stop formant transition constraints are not strengthened so much (say, through a Hebbian mechanism). In contrast, the inputs they confront everyday all conform to the nasal allophonic constraints, which are hence strengthened much more.

5 Conclusion

If the finding is real and the speculation is on the right track, the results obtained in this paper indicates that constraints have to be "trained." Training is a matter of degree, and hence constraints are weighted (have different strengths). In other words, human linguistic knowledge, conceived as a set of constraints, is not simply a set of constraints with an equal status (a non-OT perspective) or a discretely ranked constraints (an OT perspective), but rather (or also) a set of numerically weighted constraints.[6]

---

(5) Keiichi Tajima (p.c.) also pointed this out.

(6) Mathematically, numerical weights could implicate discrete ranking. However, I have not investigated whether the ranking in the OT sense should be subsumed under this picture.

References

Fromkin, Victoria A. (ed.) (2000) *Linguistics: An Introduction to Linguistic Theory*. Malden, MA, and London, UK: Blackwell.

Kakehi, Kazuhiko, Kazumi Kato, and Makio Kashino (1996) "Phoneme/Syllable Perception and the Temporal Structure of Speech." In Takashi Otake and Ann Cutler (eds.) *Phonological Structure and Language Processing: Cross-Linguistic Studies*. Berlin and New York: Mouton de Gruyter, pp. 125–143.