Phonemic and Featural Modeling of New L2 Sound Acquisition

Kakeru Yazawa[1] & Paola Escudero[2]

[1]*Waseda University (Japan),* [2]*Western Sydney University (Australia)*
k-yazawa@aoni.waseda.jp, paola.escudero@westernsydney.edu.au

It is widely accepted that nonnative speech perception patterns are predictable from phonetic and phonological similarities between the first (L1) and second (L2) language sound categories [1]. Previous L2 studies and models therefore often assume that phonemes are the most basic unit, but emerging evidence suggests that phonological features constituting the phonemes play an important role [2, 3]. This study conducts two kinds of computational modeling, phonemic and featural, which provide different kinds of explanations regarding how a new L2 sound is acquired.

Escudero and Boersma [4] proposed that perception of sound categories can be modeled by Optimality-Theoretic (OT) constraints such as "a value of *x* on the auditory continuum *y* should not be perceived as the phonological category *z*" (e.g., "[F1 = 800 mel] is not /a/"). Tableau 1 illustrates how such constraints can express the perception of a typical token of /a/ with the acoustic values of [F1 = 800 mel] and [F2 = 1100 mel]. In this example, the constraint "[F2 = 1100 mel] is not /e/" is ranked the highest, perhaps because an F2 of 1100 mel is too back for /e/. Thus, the alternative candidate /a/ is chosen as the winner ('perceived'). Escudero [1] argues that this kind of acoustic-to-phoneme modeling adequately explains perception.

In contrast, Boersma and Chládková [5] argue that acoustic-to-feature modeling better explains speech perception. They proposed that a sound category (e.g. /e/) should be expressed as a bundle of phonological features (e.g. /mid, front/). Feature co-occurrence is prohibited by constraints such as "*/mid, central/" whose strictness depends on the organization of features in the particular language [6]. Tableau 2 shows the perception of [F1 = 800 mel, F2 = 1100 mel] (same as Tableau 1) as /low, central/ (i.e. /a/). Here, the constraints "*/low, front/" and "*/mid, central/" are ranked high, perhaps because these combinations of features do not occur in this language. The third constraint "[F2 = 1100 mel] is not /front/" then prohibits /mid, front/, and thus /low, central/ is chosen as the winner.

I will now apply the two approaches to a real L2 learning scenario, namely L1 Japanese listeners' perceptual acquisition of L2 American English (AmE) vowels /ɛ, æ, ʌ, ɑ/. According to Strange et al. [7], these vowels are spectrally assimilated to either /e/ or /a/ in Japanese. The learner's task is to acquire new phonemes, or new sets of features, out of the existing ones to deal with the L2 listening environment. Phonemic and featural simulations were performed based on Stochastic Optimality Theory and the Gradual Learning Algorithm [8], using real acoustic values [9].

Figs. 1 and 2 show the results of L1 Japanese and L2 AmE simulations, respectively. Phonemic and featural modeling yielded very similar outcomes, but there was a crucial difference in how these models were implemented. First, the virtual phonemic listener learned to perceive two categories (/e, a/) in their L1 because these are the only possible categories, which they transferred to L2 perception. In order for new L2 category formation to take place, the model needed to be explicitly told that there were four categories (/ɛ, æ, ʌ, ɑ/) instead of two in the L2. This equals introducing new candidates (e.g. /æ/) and constraints related to the candidates (e.g. "[F1 = 800 Hz] is not /æ/") in terms of OT modeling.

On the other hand, the featural listener learned to perceive /mid, front/ (i.e. /e/) and /low, central/ (i.e. /a/) in their L1, even though perception of the other feature combinations (/mid, central/ and /low, front/) was also possible in principle. This is because constraints against infrequent feature co-occurrences are ranked high, while those against frequent co-occurrences are ranked low. As the virtual listener got exposed to L2 AmE in which /mid, central/ (i.e. /ʌ/) and /low, front/ (i.e. /æ/) do occur unlike the L1, they gradually lowered the ranking of "*/mid, central/" and "*/low, front/." The resultant model could perceive all four possible combinations of features: /mid, front/ (/ɛ/), /mid, central/ (/ʌ/), /low, front/ (/æ/) and /low, central/ (/ɑ/).

Although both phonemic and featural modeling can express the acquisition of new L2 sounds, featural modeling may provide an ecologically more valid account. The strength of featural modeling is that a new ‘category’ can emerge based on re-used L1 features, unlike phonemic modeling where new categories have to be intentionally added. Featural modeling may also explain why certain L2 sounds are perceived as poor exemplars of an L1 sound. For example, the featural L1 Japanese listener is aware that AmE /æ/ is /low/ and /front/, but nonetheless perceives it as /mid, front/ (/e/) or /low, central/ (/a/) because the /low/ and /front/ features do not occur simultaneously in Japanese.

Tableau 1 Perception of [F1 = 800 mel, F2 = 1100 mel] as /a/.

[F1=800 mel, F2=1100 mel]	[F2=1100] not /e/	[F1=800] not /e/	[F1=800] not /a/	[F2=1100] not /a/
/e/	*!	*		
☞/a/			*	*

Tableau 2. Perception of [F1 = 800 mel, F2 = 1100 mel] as /low, central/.

[F1=800 mel, F2=1100 mel]	*/low, front/	*/mid, central/	[F2=1100] not /front/	[F1=800] not /mid/	*/low, central/
/mid, front/			*!	*	
/mid, central/		*!		*	
/low, front/	*!		*		
☞/low, central/					*

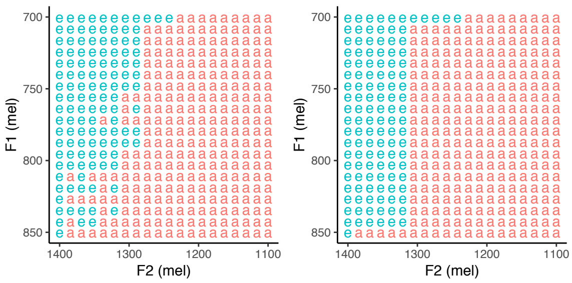

Fig.1 Phonemic (left) and featural (right) simulations of L1 Japanese perception.

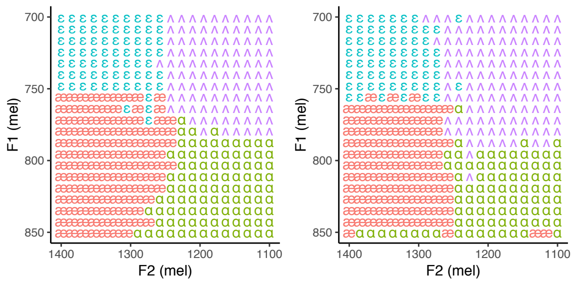

Fig.2 Phonemic (left) and featural (right) simulations L2 AmE perception.

References

[1] Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* (Doctoral dissertation). Utrecht University.

[2] Balas, A. (2017). More evidence for selective attention to features: Perception of the Turkish high back unrounded vowel. In *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2017* (pp. 65–70).

[3] Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science*, *343*(6174), 1006–1010.

[4] Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*(4), 551–585.

[5] Boersma, P., & Chládková, K. (2011). Asymmetries Between Speech Perception and Production Reveal Phonological Structure. In *Proceedings of the 17th International Congress of Phonetic Science* (pp. 328–331).

[6] Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1013–1016).

[7] Strange, W., Akahane-Yamada, R., Kubo, R., Trent-Brown, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, *26*(4), 311–344.

[8] Boersma, P. (1998). *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives* (Doctoral dissertation.) University of Amsterdam.

[9] Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *The Journal of the Acoustical Society of America*, *124*(1), 576–588.