

PERCEPTION, PRODUCTION, AND PERCEPTION–PRODUCTION

Research findings and implications for language pedagogy

By Charles Nagle, Iowa State University

When we are born our perceptual systems are capable of discriminating sounds that occur in English, Spanish, Hindi, or any other language. During the first year, our perception begins to zero in on the particular set of sounds that are contrastive in our native language(s) (L1s) (Kuhl et al., 2006). For example, a child whose parents are L1 English speakers will pick up on the fact that /b/ and /p/ are contrastive in English (e.g., “bet” vs. “pet”) and that the major difference is in the burst of air that occurs when the stop is released (i.e., there is a stronger burst of air, or more aspiration, on /p/ than /b/). A child whose parents are L1 Hindi speakers will pick up on this contrast, which also occurs in Hindi, as well as other contrasts that occur in Hindi but not in English. As our perception becomes attuned to our L1(s), we become more sensitive to L1 contrasts, such as /b/ vs. /p/ for L1 English speakers, and less sensitive to non-native contrasts, even though our ability to discriminate non-native sounds remains intact. When we begin to learn another language (L2) later in life, be it through formal instruction at university or through immersion if we move to another country where a different language is spoken, our L1 acts as a filter, altering our perception of L2 sounds. Consequently, we may not detect differences between contrastive L2 sounds that are not contrastive in our L1, and we may fail to notice the difference between our accented pronunciation of the L2 and the target pronunciation. For example, an English speaker who is learning L2 Hindi would probably perceive dental and retroflex stop consonants /t̪/ and /t̠/, which are contrastive in Hindi but do not occur in English, as variants of English alveolar stops /t/ and /d/. According to major theories of L2 pronunciation learning such as the Speech Learning Model (Flege, 1995, 2003), if we do not perceive differences between similar L1 and L2 sounds, then we will not produce the corresponding L2 sounds accurately. In other words, accurate perception is a necessary condition for consistent accurate production. At the same time, perception and production involve distinct cognitive and motor skills, so development across the two modalities may not be synchronized.

Key terminology

The following tasks are typically used to assess perception:

1. **Identification:** Hear a word and select the written word or image to which it corresponds.
2. **Discrimination:** In an AX task, hear an anchor word (A) and another word (X) and decide if they are the same; in an ABX task, hear two anchor words (A and B) and decide if a third word (X) corresponds to A or B.
3. **Oddity or oddball:** Hear three words, decide if there is an odd word out, and indicate the position of the odd word (1, 2, or 3). If there is no odd word, indicate that all words are the same.

The following tasks are typically used to assess production:

1. **Word & sentence reading:** See a word or sentence and read it aloud.
2. **Word & sentence repetition:** Hear a word or sentence and repeat it.
3. **Picture description:** See a picture and describe it in a few sentences.
4. **Picture narration:** See a series of images that tell a story and narrate the story.

Are perception and production related?

Research has shown that perception and production accuracy are related, though the strength of the relationship may vary depending on the proficiency of the speaker-listener and the target sounds (Flege, Bohn, & Jang, 1997; Flege, MacKay, & Meador, 1999; Saito & van Poeteren, 2017). For instance, Saito and van Poeteren studied L1 Japanese speakers' perception and production of the English /l/-/ɹ/ contrast. Japanese speakers typically struggle with this contrast because they perceive these two English sounds as instances of a single Japanese “r” category (an alveolar tap or flap /ɾ/). Saito and van Poeteren assessed perception using an identification task (hear a word containing /l/ or /ɹ/ and select the correct word from two written options, such as “rink” vs. “link”) and production through reading and picture description. Production accuracy was defined in terms of acoustic measurements and listener perception. For the latter, native English speakers evaluated the quality of learners' /ɹ/ production using a nine-point scale (1 = “very good /ɹ/”, 5 = “neither /ɹ/ nor /l/”, and 9 = “very good /l/”), which the authors also recoded into an intelligibility judgment (i.e., scores of 1–4 corresponding to the /ɹ/ portion of the continuum were deemed intelligible). Perception accuracy was correlated with the impressionistic production measures, but results were more variable for the acoustic measurements. These findings suggest that perception was more closely aligned with the production of intelligible L2 sounds than with the production of native-like acoustic characteristics (e.g., the use of F3¹).

Does the relationship between perception and production change over time?

If accurate perception facilitates accurate production, then the characteristics of the link itself deserve our attention. In other words, exactly what type of relationship is evident between perception and production? One possibility is that the two modalities develop in tandem. Although this view is intuitively appealing, longitudinal studies tracking perception and production in the same sample of learners over time paint a more complex picture. For example, Hanulíková, Dediu, Fang, Bašňáková, and Huettig (2012) trained multilingual L1 Dutch speakers on Slovak consonant clusters such as /vzbl:knuc/ (“to burst”). Over three sessions, participants completed a range of perception measures designed to tap into different skills (e.g., mispronunciation detection), and production was assessed through word reading and word imitation. Participants' production accuracy was subsequently evaluated by native Slovak speakers using a seven-point scale with higher scores indicating better performance. Despite substantial individual variation in performance across all tasks, only mispronunciation detection and word reading were related to one another. On

¹ F1, F2, and F3 are acoustic measures that refer to resonances in the vocal tract. F1 and F2 are similar in English /ɹ/ and /l/, but F3 occurs at a much lower frequency in /ɹ/. Native English speakers predominantly rely on F3 when discriminating the two sounds (e.g., *rip* vs. *lip*), but L1 Japanese speakers typically struggle to perceive this acoustic cue.

the basis of these findings, the authors hypothesized that perception and production may dissociate during the early stages of learning or that production might only improve once more accurate perceptual representations form.

As illustrated in Figure 1, three characterizations of the perception–production link are possible: (1) a synchronous relationship: gains in perception feed into production relatively quickly and efficiently; (2) a threshold effect: production begins to improve once perception reaches a certain level of accuracy, as Hanulíková et al. (2012) discussed; and (3) a lagged relationship: perception and production follow similar trajectories but production begins to improve at a slightly later stage.

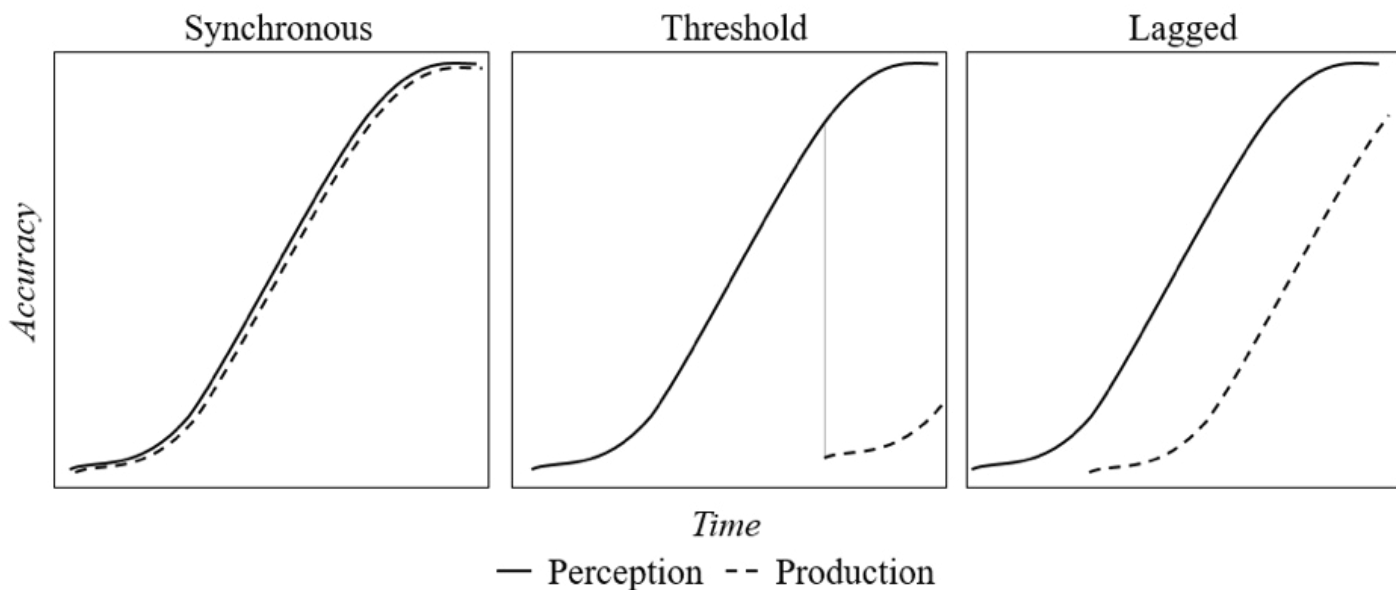


Figure 1. Possible relationships between perception and production.

I tested these different models by investigating how L1 English speakers perceive and produce L2 Spanish stop consonants (Nagle, 2018). As noted above, English voiced and voiceless stops (e.g., /b/ vs. /p/) are differentiated by degree of aspiration: voiceless stops such as /p/ are produced with a stronger burst of air (compare “pat” and “bat” by placing your hand in front of your mouth as you pronounce both words). Put another way, English contrasts unaspirated voiced stops and aspirated voiceless stops. In Spanish, voiced stops are voiced (there is no burst of air) and voiceless stops are produced with a very short burst similar to English /b/. The Spanish system could therefore be described as a voicing/unaspirated contrast. English speakers need to recalibrate their perception to fit the voicing/unaspirated distinction in Spanish, and to improve their production, they need to produce full voicing in voiced stops and reduce aspiration of voiceless stops. Participants completed an identification task and a semi-controlled sentence production task five times while enrolled in second, third, and fourth semester university-level Spanish language courses. Analyses suggested a lagged relationship between the perception and production of Spanish /p/, insofar as gains in identification accuracy were correlated with increased

production accuracy at the subsequent session. However, no relationship was evident between the perception and production of /b/, which could be due to the fact that fully voiced stops are challenging to produce.

Overall, the results of these studies indicate that perception may benefit production when the learning task involves altering a familiar setting such as reducing aspiration of voiceless stops in the case of L1 English/L2 Spanish. In contrast, accurate perception may not be enough to help learners master a new sound whose characteristics are fundamentally distinct from those used in the L1. In terms of the nature of the perception–production link itself, it seems probable that perception will lead production and that targeted training may be needed to boost the latter in some cases.

Does perception training enhance production and vice versa?

In a recent meta-analysis synthesizing findings from 18 studies, Sakai and Moorman (2017) found that perception training leads to medium gains in perception and small but reliable gains in production, especially for obstruents (e.g., stop consonants like /b/ and /p/ and fricatives like /s/ and /z/). Sakai and Moorman furthermore found that certain training characteristics such as a short interval of 3.5 hours or less seem to facilitate production gains. In addition to these variables, recent work demonstrates that sleep may play an important role in phonetic learning. Earle and Myers (2015) trained L1 English listeners on Hindi dental and retroflex stops, which English speakers typically perceive as instances of English alveolar /d/. In a series of experiments, the authors manipulated the timing of the training (i.e., morning vs. evening) and L1 exposure (i.e., exposure to English /b/ or /d/). Listeners' identification of Hindi stops was relatively stable, but their discrimination of the contrast was susceptible to sleep and L1 interference. In particular, discrimination improved after sleep, but if learners were exposed to L1 /d/ tokens after the training but prior to sleep, discrimination did not improve overnight. Production training can also lead to gains in perception. Kartushina, Hervais-Adelman, Frauenfelder, and Golestani (2015) trained L1 French learners on L2 Danish vowels over five sessions, assessing pre- and post-training perception using a discrimination task and production using word repetition. Highlighting perception–production asymmetries, pretraining testing revealed that the vowels that were more challenging to perceive were not necessarily more challenging to produce. Nevertheless, the group that received visual feedback training improved their perception and production of the Danish vowels, and the correlation between gains in perception and production was nearly significant.

Though training in one modality can lead to gains in the other, asking learners to produce sounds during perception training may actually destabilize the emerging perceptual system. In Baese-Berk and Samuel (2016), participants were randomly assigned to a perception only group or a perception and production group. Both groups were trained on an L2 Basque contrast using an ABX discrimination task (hear stimuli A and B and decide

if a third stimulus, X, is more similar to A or B), but the perception and production had to repeat the final token (X) before making the perceptual judgment. In a follow-up study, the combined group read a random letter aloud rather than repeating the target stimulus. The perception only group displayed sensitivity to the contrast, but the combined training group did not, irrespective of whether they repeated the final token or read a letter aloud. This finding indicates that any type of speech production, including production that is not related to the target contrast, can compromise perceptual learning.

How might individual differences affect the perception–production link?

Though individual difference research typically focuses on one modality or the other, we can also imagine how relationships between aptitude and learning (Bowles, Chang, & Karuzis, 2016) or between (in)accurate self-perception and production (Trofimovich, Isaacs, Kennedy, Saito, & Crowther, 2014) might influence the extent to which perception and production are interconnected. For example, individuals who possess an aptitude for distinguishing subtle variations in L2 sounds and encoding those differences might learn to discriminate L2 contrasts more quickly, be it in the language classroom or through explicit training. This, in turn, could set the stage for more rapid gains in production. At the same time, individuals who are able to analyze their own speech and recognize differences between their production and the L2 target may be able to align their perception and production more quickly. At this stage, these relationships are speculative since research has yet to investigate how individual differences in aptitude, self-assessment, and language use shape perception–production when the link is construed as an integrated process. However, these relationships could help explain the substantial variation that is evident in both group-level and individual analyses of the perception–production link.

Summary and take-aways

Research has generally yielded the following set of findings for the perception–production link:

- Accurate perception is a necessary, but in some cases insufficient, condition for accurate production.
- Perception and production are related, but the strength and form of that relationship may vary over time and as a function of the target structure; shared features (features that occur in both the L1 and L2) may exhibit a closer connection than novel features (features that occur only in the L2), which may be more difficult to acquire.
- In certain contexts of learning (perhaps especially in an instructed or classroom context), perception and production will probably change at different rates, with production improving after perception.
- Training one modality can lead to gains in the other, but engaging the production

system during perception training can compromise emerging perceptual representations.

- Sleep seems to enhance perceptual learning, particularly if training occurs immediately beforehand (i.e., limited exposure to the L1 between training and sleep).
- Individual differences associated with more accurate perception and production may also influence the perception–production link.

Drawing upon these findings, we can begin to envision a pedagogical approach that could maximize gains in both modalities and potentially facilitate a closer connection between them. The following example deals with L2 English vowels.

Week 1

Evaluate students' ability to perceive (discriminate and identify) and produce vowels to determine shared and individual needs.

Discuss the characteristics of English vowels, drawing students' attention to important minimal pairs.

Week 2

Drawing upon the results of the vowel analysis, design and implement identification and discrimination tasks involving minimal pairs.

Introduce students to high variability phonetic training available through [English Accent Coach](#) (Thomson, 2018) and assign students to complete 30 minutes of training three times per week in the evening, ideally shortly before sleep.

Week 3

Reassess vowel perception and continue perception training as needed inside and outside of class. In all likelihood, multiple weeks of training will be required.

Week 4

Design and implement controlled production activities focusing on the vowels that students struggled to produce, emphasizing intelligibility over nativelike accuracy (Levis, 2005).

Design and implement information gap activities such as map tasks, spot the difference, etc. (for an example, Solon, Long, & Gurzynski-Weiss, 2016) to practice vowels within a communicative framework.

Despite the complexities of the perception–production link, we should not lose sight of the fact that pronunciation instruction is effective (Lee, Jang, & Plonsky, 2014; Thomson & Derwing, 2014). Ultimately, we must strive to take a principled approach to perception and production and to integrate training into our teaching on a systematic basis.

References

- Baese-Berk, M. M. & Samuel, A. G. (2016). Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language*, 89, 23–36.
<https://doi.org/10.1016/j.jml.2015.10.008>
- Bowles, A. R., Chang, C. B., & Karuzis, V. P. (2016). Pitch ability as an aptitude for tone learning. *Language Learning*, 66(4), 774–808. <https://doi.org/10.1111/lang.12159>
- Earle, F. S. & Myers, E. B. (2015). Sleep and native language interference affect non-native speech sound learning. *Journal of Experimental Psychology: Human Perception and Performance*, 41(6), 1680–1695. <https://doi.org/10.1037/xhp0000113>
- Flege, J. E. (1995). Second language speech learning: Theory, findings, problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). Timonium, MD: York Press.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319–355). Mouton de Gruyter: New York, NY.
- Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.
<https://doi.org/10.1006/jpho.1997.0052>
- Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 106(5), 2973–2987. <https://doi.org/10.1121/1.428116>
- Hanulíková, A., Dediu, D., Fang, Z., Bašňaková, J., & Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: A training study. *Language Learning*, 62(S2), 79–109.
<https://doi.org/10.1111/j.1467-9922.2012.00707.x>
- Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *Journal of the Acoustical Society of America*, 138(2), 817–832.
<https://doi.org/10.1121/1.4926561>
- Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2), F13–F21. <https://doi.org/10.1111/j.1467-7687.2006.00468.x>
- Lee, J., Jang, J., & Plonsky, L. (2014). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, 36(3), 345–366.
<https://doi.org/10.1093/applin/amu040>
- Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588485>
- Nagle, C. (2018). Examining the temporal structure of the perception–production link in second language acquisition: A longitudinal study. *Language Learning*, 68(1), 234–270.
<https://doi.org/10.1111/lang.12275>
- Saito, K. & van Poeteren, K. (2017). The perception–production link revisited: The case of Japanese learners' English /ɹ/ performance. *International Journal of Applied Linguistics*.
<https://doi.org/10.1111/ijal.12175>

- Sakai, M. & Moorman, C. (2017). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187–224. <https://doi.org/10.1017/S0142716417000418>
- Solon, M., Long, A. Y., & Gurzynski-Weiss, L. (2016). Task complexity, language-related episodes, and production of L2 Spanish vowels. *Studies in Second Language Acquisition*, 39(2), 347–380. <https://doi.org/10.1017/S0272263116000425>
- Thomson, R. I. (2018). *English accent coach* (Version 2.3). Retrieved from <https://www.englishaccentcoach.com/>
- Thomson, R. I. & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36(3), 326–344. <https://doi.org/10.1093/applin/amu076>
- Trofimovich, P., Isaacs, T., Kennedy, S., Saito, K., & Crowther, D. (2014). Flawed self-assessment: Investigating self- and other-perception of second language speech. *Bilingualism: Language and Cognition*, 19(01), 122–140. <https://doi.org/10.1017/S1366728914000832>

Author Bio



Charles Nagle is an Assistant Professor of Spanish Linguistics and the Director of the Spanish Language Program at Iowa State University. His research focuses on L2 pronunciation and individual differences, particularly how learners' pronunciation develops over time and the relationship between the perception and production of second language sounds. He is also interested in teachers' beliefs on pronunciation learning and teaching.