Second Language Speech Perception

DANIELLE DAIDONE

What Is Speech Perception?

Speech perception broadly refers to the process of interpreting the sounds of spoken language. When listening to language, listeners decipher which speech sounds are present in the auditory signal and match them to representations stored in memory, which are typically referred to as "words" but can also correspond to any unit that has an independent function in language use, from idioms such as *break a leg* to smaller units such as *-ness*. Successful listeners must be able to correctly interpret this spoken input despite the speed, continuity, and variability of speech. To exemplify the speed of spoken language, Cutler (2012) reported that a 10-second excerpt of an American English speaker talking at a comfortably slow rate consisted of 50 words. This means that a listener would have to process 6.8 words per second to understand this relatively slow speech sample. These words together consisted of 175 vowels and consonants, which equals 17.5 different sounds that must be processed per second (pp. 33–4). Not only is spoken language fast, but it is also not conveniently divided into words. Words typically do not have pauses between them within the same clause, and silence does not always correspond to a pause between words. For example, when speakers pronounce the /p/ in the word *capital*, their lips come together for a moment, which prevents air from leaving the mouth. This creates a brief period of silence during the /p/, but this silence comes in the middle of a word rather than between words. Therefore, listeners cannot rely on periods of silence to divide the spoken language they hear into words. Furthermore, speech is variable, even beyond the adjustments that must be made in order to understand different voices. The word *fishing* may be pronounced as *fishin'* ['fɪ.ʃɪn] or *fishing* ['fɪ.ʃɪŋ], *going to* may be pronounced as *going to* ['ɡoʊ.ɪŋ tʰu] or *gonna* ['ɡʌ.nə] and the first syllable in *either* may be pronounced with the vowel in *eat* ([i]) or with the vowel in *eye* ([aɪ]). Variability in speech is due to many different factors: these can be linguistic, such as the characteristics of the surrounding sounds or the frequency of a word, or extralinguistic, such as the speaker's age, the intended audience, or the acoustics of the room. Many different factors simultaneously affect the speech signal, making variation ubiquitous. Yet despite the fast, continuous, and variable nature of spoken language, speech perception is typically automatic and effortless for individuals listening to their native language.

What Difficulties Can Learners Have with Speech Perception in a Second Language?

Speech perception is often less successful for second language (L2) learners, who can have difficulties at any stage in the processing of spoken language. At the lowest level of processing, non-native listeners can struggle with decoding the auditory input into phonetic categories. L2 misperceptions can be broadly discussed in terms of four partially overlapping types. First of all, learners may not perceive a distinction between L2 sounds. A well-known example is Japanese speakers' difficulty in telling the difference between English /l/ and /ɹ/, as in *light* versus *right*. Another demonstration of this type of difficulty is Spanish speakers' struggle with the English vowel contrast /ʌ/ (e.g., *cut*)

The Encyclopedia of Applied Linguistics. Edited by Carol A. Chapelle
(Section Editors: John Levis, Murray Munro, Charles Nagle and Amanda Huensch).

DOI: 10.1002/9781405198431.wbeal20649

versus /æ/ (e.g., *cat*), as documented by learners' relatively low performance in an experiment requiring them to differentiate these sounds (Baigorri et al., 2019). Second, learners may misperceive one L2 sound as another. For instance, Hallé and Best (2007) showed that English and French speakers tend to perceive initial /tl/ clusters in Hebrew as /kl/ instead. Third, learners may perceptually delete an L2 sound. Wang et al. (2023) found that Mandarin-speaking learners of English were frequently unable to distinguish stimuli with and without an /l/ after certain vowels, illustrated by words such as *howl* and *how*. Finally, L2 learners may perceptually insert a sound. For example, a Spanish speaker may not only pronounce *spy* as *espy*, with an initial [e], but perceive it this way as well. Insertion of this kind of "epenthetic" vowel has been investigated by researchers such as Dupoux and colleagues (2011), who found that speakers of Brazilian Portuguese reported hearing /i/ between the consonants in stimuli like *ebzo*, while speakers of Japanese reported hearing /u/. In sum, misperceptions can take the form of a lack of perceptual distinction between two L2 sounds, the perception of one sound or sequence of sounds as another, the deletion or failure to detect a sound present in the acoustic signal, or the insertion of a sound not present in the signal.

Difficulties with decoding the speech stream can also occur beyond the level of individual sounds, exemplified by features such as vowel and consonant length, syllable stress, and tone. One example is the perception of stress by French-speaking learners of Spanish, who tend to struggle with differences in stress placement that can differentiate words like *hablo* /'a.blo/ (with initial stress) "I speak" and *habló* /a.'blo/ (with final stress) "s/he spoke" (Dupoux et al., 2008).

The strategies employed to segment speech, or find the boundaries between words, are also frequently less efficient for L2 learners than for first language (L1) speakers. One cue to the location of word boundaries is whether combinations of sounds in the speech stream are possible at the beginning and ends of words (i.e., phonotactics). For example, an English word can begin with /sl/ as in *sleep* but a German word cannot, and thus in German if a listener hears the sequence /sl/ they know it must span across a word boundary, such as in *das Lamm* "the lamb." In an experiment by Weber and Cutler (2006), German learners of English experienced interference from German word boundary constraints, and thus were faster at recognizing *lecture* when it was preceded by /s/, whereas L1 English speakers were not.

Another difficulty learners face comes at the level of processing in which sounds are matched to representations of words stored in long-term memory. If learners struggle to differentiate L2 sounds, this can cause them to activate the wrong word in memory (e.g., activating *fly* when hearing *fry*). This can happen even when words only partially overlap; Cutler et al. (2006) found that Japanese learners of English temporarily activated *locker* when hearing *rocket*. In addition, accurate perception does not guarantee accurate word retrieval. Díaz et al. (2012) found that almost half the Dutch-speaking participants in their study scored within the native range for a task in which they categorized sounds as either the English vowel /æ/ or /ɛ/; however, only a few participants scored within the native range for tasks that required them to identify words containing those same vowels, as in *bat* and *bet*, respectively.

The way in which L2 learners have stored the sounds of words in long-term memory (i.e., their "phonolexical representations") can also be imprecise or inaccurate. Misperceptions of L2 sounds may form part of phonolexical representations, but difficulties can also occur for words that do not contain new L2 sounds or sound combinations. For example, Farrell and Daidone (2023) found that L1 English learners of Spanish often believed that Spanish words spelled with <h> contained an [h] sound, despite <h> not being pronounced in Spanish. Overall, research has demonstrated that learners can experience a variety of problems with L2 perception and speech processing.

Why Is Second Language Speech Perception Difficult?

Studies examining speech perception by infants have established that the ability to accurately differentiate non-native sounds diminishes very early; non-native vowels can become difficult

to distinguish by 6–8 months and consonants by 10–12 months (see Polka & Werker, 1994). As explained by the Native Language Magnet Model, subsequently updated as the Native Language Magnet Theory, Expanded (Kuhl, 1994; Kuhl et al., 2008), experience with the native language warps infants' perception. This means that infants show increased attention to dimensions of the speech signal that are relevant for their L1 and less sensitivity to dimensions that are not, which is reflected in the establishment of neural networks optimized for L1 perception. The optimization of attention to those acoustic cues that are relevant in the L1 creates a perceptual magnet effect that minimizes perceived differences near the prototype of a native sound category and maximizes the perceived differences at the boundaries of categories. This perceptual warping aids in L1 language development. Kuhl and colleagues (2008) demonstrated that infants who have more optimized L1 perception, and thus less accurate non-native perception, have subsequent faster L1 linguistic development than those who retain more accurate non-native perception. However, this L1 optimization in infancy also makes learning an additional language later on in life more difficult since the L2 will typically rely on different acoustic cues or use them in a different way. Importantly, it is not listeners' basic ability to perceive sensory input that has changed, but rather their attunement to the signals in it for linguistic purposes. For example, English speakers are capable of hearing differences in pitch, but English does not use pitch to differentiate words like in Mandarin Chinese, and thus English-speaking learners of Chinese struggle to acquire words that differ by tone, such as *má* 麻 (with a rising tone) "hemp" and *mǎ* 马 (with a dipping tone) "horse."

Given the importance of the L1 in shaping L2 perception, various accounts have been proposed to explain how the relationship between native and non-native sounds affects their perception and acquisition. The Speech Learning Model (SLM), developed by Flege (1995) and later updated by Flege and Bohn (2021) as the revised Speech Learning Model (SLM-r), focuses on the development of the L2 phonological system for both the perception and production of sounds. If a learner cannot differentiate between an L2 sound and its closest L1 counterpart, then these sounds will be linked in a single category, and over time the production of this linked category will reflect both the L1 and L2 sounds it represents. For example, Flege (1991) reported that learners who have shared categories for L1 and L2 stop consonants, such as Spanish /t/ and English /t/, exhibit a compromise pronunciation that falls between the average values of the L1 sound and the L2 sound. Alternately, depending on properties of the sounds and the amount of L2 experience, learners may begin to notice some of the phonetic differences between the L1 and L2 sounds and create a separate L2 category.

The Perceptual Assimilation Model (PAM) (Best, 1995), subsequently extended to address L2 learning (PAM-L2; Best & Tyler, 2007), focuses on the perception of contrasts rather than individual sounds. According to this model, if two L2 sounds are perceived as close to two different L1 categories (*two-category assimilation*), they are expected to be easy to discriminate from each other. However, if two L2 sounds are perceived as similar to the same L1 sound (*single-category assimilation*), these L2 sounds are predicted to be difficult to differentiate, and it is unlikely that a learner will form two new L2 categories for these sounds. It may also be that one of the two L2 sounds is perceived as very similar to that L1 sound, while the other L2 sound is more noticeably different from it (*category-goodness assimilation*). L2 sounds exhibiting a category-goodness assimilation are predicted to be easier to tell apart than those in a single-category assimilation, and it is expected that learners will form a new category for the more distinct L2 sound. Another assimilation possibility is that L2 sounds are not perceived as particularly close to any L1 sounds; these sounds are termed *uncategorized*. The final possibility in PAM-L2 is that L2 sounds are not perceived as speech, and therefore cannot be assimilated to L1 categories (*non-assimilable*). An instance of this assimilation type would be American English speakers listening to Zulu click consonants. According to Best and Tyler, it is unclear whether these types of unassimilable sounds can ever be acquired as new L2 categories.

Other competing models also exist in the field, such as the Second Language Linguistic Perception model (L2LP) (Escudero, 2005; van Leussen & Escudero, 2015), which focuses on acoustic

differences between L1 and L2 sounds to make predictions about acquisition, and the Automatic Selection Perception (ASP) model (Strange, 2011), which emphasizes the importance of selective perception routines. While all of these models differ in their specifics, they share the tenet that the existing L1 sound system shapes how the L2 is perceived. L2 phonological acquisition is proposed to depend on the various relationships between L1 and L2 acoustic cues, sounds, or contrasts, making the learning of some parts of the L2 sound system more difficult than others.

How Does Second Language Speech Perception Improve?

As described by Flege and Bohn (2021) in SLM-r, the quantity and quality of L2 input affects the acquisition of L2 sounds, since input serves as evidence for the distribution of L2 sounds and the cues that differentiate them. While past studies have frequently used length of residence (LOR) as a proxy for the quantity of L2 input, Flege and Bohn propose the metric of full-time equivalent (FTE) years of L2 input, or the LOR in a predominantly L2 environment multiplied by the proportion of L2 use during that time. Researchers generally agree that the more time learners spend interacting in the L2, the more robust their L2 speech perception is likely to be, but there remains debate about the effects of input and the best way to measure input quality and quantity (Gorba, 2023).

L2 experience tends to be related to L2 vocabulary size, which has also been shown to be related to perception ability. The expansion of a learner's L2 vocabulary is hypothesized to encourage the development of their L2 phonological system; as learners acquire more minimally different words, this puts pressure on their phonological system to pay attention to L2 specific cues to better differentiate the L2 words. In a study examining Japanese learners of English, Bundgaard-Nielsen et al. (2011) found that those with a larger vocabulary size had more accurate perception of English vowel contrasts compared to those with a smaller vocabulary. Vocabulary size has also been shown to predict how accurately learners store the sounds of words in memory (Daidone & Darcy, 2021). Collectively, studies in this area suggest that acquiring more L2 words can lead to better L2 speech processing.

Research additionally shows that L2 learners can improve their L2 speech perception through high variability phonetic training (HVPT) (see Thomson, 2018). In HVPT, learners listen to speech sounds and are asked to discriminate them or identify them. These sounds are presented in a variety of voices and phonetic contexts, which helps learners develop the ability to reliably recognize them despite the inherent variability of speech. One of the pioneering examples of HVPT is a series of studies training native Japanese speakers to perceive English /l/ and /ɹ/ (Lively et al., 1993). In these studies, Japanese participants listened to English words produced by various talkers and had to choose the correct response from two possible options. For example, if participants heard a word like *fry*, they would choose between the options *fly* and *fry* displayed on the screen. They were then given feedback on whether their response was correct or not, and the stimulus repeated if they made an error. These researchers found that listeners improved their ability to identify English words differing in /l/ and /ɹ/ and were able to generalize to unfamiliar talkers. Importantly, those listeners exposed to only one talker instead of five were not able to generalize to words spoken in a new voice. The authors argue that the HVPT paradigm enables listeners to adjust their attention to focus on the acoustic cues relevant for contrasts in a non-native language. A later study by Francis et al. (2000) confirmed that phonetic training can lead to changes in the cues used by listeners.

Conclusion

While native speakers of a language typically do not have problems with processing speech despite its rapid, continuous, and variable nature, L2 learners can exhibit various difficulties. The L1 profoundly shapes the perception and acquisition of a non-native phonology, and thus learners may

struggle to appropriately interpret L2 sounds, leading to problems such as perceptual illusions and misinterpretations. Nevertheless, learners are able to improve their L2 speech perception through measures such as increasing their L2 input and vocabulary size and engaging in perceptual training tasks.

SEE ALSO: Bilingualism and Speech Perception; Comparative Phonetics and Phonology; Phonological Acquisition; Pronunciation Teaching Methods and Techniques; Second Language Speech Perception and the Brain; Social Influences on Second Language Speech Acquisition; Spoken Word Recognition in Bilingualism

References

Baigorri, M., Campanelli, L., & Levy, E. S. (2019). Perception of American-English vowels by early and late Spanish-English bilinguals. *Language and Speech*, *62*(4), 681–700. https://doi.org/10.1177/0023830918806933

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, *33*(3), 433–461. https://doi.org/10.1017/S0272263111000040

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. MIT Press.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269–284. https://doi.org/10.1016/j.wocn.2005.06.002

Daidone, D., & Darcy, I. (2021). Vocabulary size is a key factor in predicting second language lexical encoding accuracy. *Frontiers in Psychology*, *12*, 688356. https://doi.org/10.3389/fpsyg.2021.688356

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680–689. https://doi.org/10.1016/j.lindif.2012.05.005

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, *64*(3), 199–210. https://doi.org/10.1016/j.jml.2010.12.004

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition*, *106*(2), 682–706. https://doi.org/10.1016/j.cognition.2007.04.001

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

Farrell, V., & Daidone, D. (2023). Effects of orthography and cognate status on second language Spanish lexical encoding. *Journal of Second Language Pronunciation*, *9*(2), 167–191. https://doi.org/10.1075/jslp.22036.far

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America*, *89*(1), 395–411.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception and Psychophysics*, *62*(8), 1668–1680. https://doi.org/10.3758/BF03212164

Gorba, C. (2023). Is full-time equivalent an appropriate measure to assess L1 and L2 perception of L2 speakers with limited L2 experience? *Languages*, *8*(1), 56. https://doi.org/10.3390/languages8010056

Hallé, P. A., & Best, C. T. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *Journal of the Acoustical Society of America*, *121*, 2899–2914. https://doi.org/10.1121/1.2534656

Kuhl, P. K. (1994). Learning and representation in speech and language. *Current Opinion in Neurobiology*, *4*(6), 812–822. https://doi.org/10.1016/0959-4388(94)90128-7

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979–1000. https://doi.org/10.1098/rstb.2007.2154

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*(3), 1242–1255. https://doi.org/10.1121/1.408177

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421–435. https://doi.org/10.1037/0096-1523.20.2.421

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*, *39*(4), 456–466. https://doi.org/10.1016/j.wocn.2010.09.001

Thomson, R. (2018). High Variability [Pronunciation] Training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231. https://doi.org/10.1075/jslp.17038.tho

van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*, 1000. https://doi.org/10.3389/fpsyg.2015.01000

Wang, Y., Bundgaard-Nielsen, R. L., Baker, B. J., & Maxwell, O. (2023). Difficulties in decoupling articulatory gestures in L2 phonemic sequences: The case of Mandarin listeners' perceptual deletion of English post-vocalic laterals. *Phonetica*, *80*(1–2), 79–115. https://doi.org/10.1515/phon-2022-0027

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, *119*(1), 597–607. https://doi.org/10.1121/1.2141003

Suggested Readings

Daidone, D., Lidster, R., & Kruger, F. (2023). Free classification as a method for investigating the perception of nonnative sounds. *Studies in Second Language Acquisition*, *28*(11), 1104–1130. https://doi.org/10.1017/S0272263123000050

Kong, E. J., & Kang, S. (2023). Individual differences in categorical judgment of L2 stops: A link to proficiency and acoustic cue-weighting. *Language and Speech*, *66*(2), 354–380. https://doi.org/10.1177/00238309221108647

Sebastián-Gallés, N. (2005). Cross-language speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 546–566). Blackwell.

Wayland, R. (Ed.). (2021). *Second language speech learning: Theoretical and empirical progress*. Cambridge University Press.