NATIVE AND NON-NATIVE SPEECH PERCEPTION

Daniel Williams[1] and Paola Escudero[2]
[1]Institute of Humanities & Creative Arts, University of Worcester, Worcester, WR2 6AJ, United Kingdom
[2]MARCS Institute, University of Western Sydney, Sydney, Australia
d_williams@worc.ac.uk

This review examines research on speech perception by both native and non-native listeners. The development of speech perception in infancy is first considered and a theoretical model that accounts for this is introduced. A brief overview then follows of several research areas under the umbrella of non-native speech perception, namely cross-dialect, cross-language and second-language speech perception. It is shown that non-native and native speech perception is critically shaped by the specific ways in which speakers use acoustic cues in speech production.

INTRODUCTION

A central goal of research on speech perception has been to uncover the auditory dimensions that listeners use to derive linguistically meaningful utterances from the speech signal. In order for listeners to successfully decode an utterance, they must be able to reliably distinguish the speech sounds of a particular language.

The speech signal presents listeners with a multitude of acoustic information, along different auditory dimensions (e.g., formants or voice onset time), that lies within the limits of human hearing. Perceiving speech is not simply a task of attending to these auditory dimensions equally; of critical importance is how listeners integrate the multiple dimensions to successfully map them onto particular speech sound categories. This crucial task is exemplified by the difficulties non-native listeners may face in discriminating non-native speech sound contrasts that make use of auditory dimensions in different ways from those in their native languages.

Consider the well-known example of the discrimination of the English /l-r/ contrast (in words such as 'lead' and 'read') by Japanese learners of English. While both native English listeners and Japanese learners attend to the same auditory dimensions, e.g., the second and third formants (F2 and F3) of /l/ and /r/, they do so in different ways. Iverson *et al*. [1] show that Japanese listeners are most sensitive to variation in F2, but not to variation in F3 which is a more reliable cue for successfully separating the two sounds. English listeners, on the other hand, are most sensitive to variation in F3 and consequently exhibit far greater discrimination accuracy.

The relative preference for certain cues over others, referred to as cue weighting, differs between native and non-native listeners. The sensitivity to these cues develops early in life and is related to the acoustic dimensions found in infants' ambient language. Unsurprisingly, Japanese does not have an alveolar sound contrast that is differentiated by F3 as in English. Though notoriously difficult, Japanese learners of English can begin to change their relative use of auditory dimensions, or attend to new dimensions, to improve their discrimination accuracy [2].

This review first provides a brief examination of how speech perception develops early in life and offers a theoretical model to account for this. The implications of this early experience are then reviewed in three related areas of non-native speech perception, namely cross-dialect, cross-language and second-language speech perception.

DEVELOPMENT OF SPEECH PERCEPTION

Theoretical accounts of native and non-native speech perception usually have in common that an individual's early experience with language shapes non-native speech perception and/or second-language learning in adulthood. Some well-known accounts include Kuhl *et al*.'s [3] Native Language Magnet model, Best's [4] Perceptual Assimilation Model and Flege's [5] Speech Learning Model.

Speech sounds are produced with great variability, yet individuals learn to identify each instance as belonging to one of a finite group of speech sound categories. Escudero's [6-8] Linguistic Perception (LP) model advocates that these categories emerge from the mapping of auditory dimensions according to how they are used and integrated in a listener's language or language variety (e.g., dialect). That is, native listeners of a particular language prefer those auditory dimensions that reliably differentiate the sounds in their speech production for that language, which is referred to as the 'optimal perception hypothesis'. Therefore listeners with different early experiences of language, and consequently divergent linguistic knowledge, will differ in how they perceive the same auditory events [6-8], as will be described below.

How might infants learn to map auditory dimensions onto speech sound categories? Evidence has repeatedly shown that infants younger than six months can discriminate most speech sounds in any language [9]. As adults, this apparent ability declines and the discrimination of non-native speech sounds becomes more difficult [10, 11, 1] (depending on the particular contrast [12]), while the discrimination of native speech sounds becomes more accurate.

Early theories on phonetic learning posited that infants possess innate phonetic capacities to distinguish all speech sounds and that the apparent decline is due to fine-tuning according to the acoustic dimensions that are relevant for sounds present in their ambient language [13, 14]. However, evidence suggests that early phonetic abilities may not be innate as some animals are capable of discrimination resembling that of humans [15] and infants are also able to discriminate non-speech sounds [16].

Infants' apparent early phonetic abilities may thus be a reflection of general auditory perception and the decline in discrimination performance in adulthood is a result of the development of speech perception, as proposed by the LP model. This perceptual development is formalised within the model as a 'perception grammar' (see [6-8] for fuller explanations of the workings of the model).

Boersma, Escudero and Hayes [17] and Escudero [8] propose that this initial perceptual development is auditory-driven and a potential mechanism is distributional learning. Instances of the same speech sound are produced with great variability on several acoustic dimensions by native speakers, but the most commonly occurring instances are those around the edges rather than the middle of an acoustic continuum. Infants have been shown to be remarkably sensitive to frequency distributions along such continua [18], ultimately leading to successful discrimination of speech sound contrasts.

The development of speech perception occurs due to changes in infants' perception grammar to cope with the distributional properties found in speech in their environment, resulting in several auditory inputs being mapped onto the most frequently perceived categories. In other words, auditory values, e.g., duration, voice onset time, first (F1) and second (F2) formants and so on, will be mapped onto a finite number of phonetic categories. Further perceptual development takes place when a lexicon develops, which within the model is referred to as lexicon-driven perceptual learning.

It is in this way that the LP model accounts for the early experiences with language in shaping speech perception in adulthood. That is, adults' perception grammars have developed for the optimal perception of instances of speech sounds as encountered in their speech environments. Hence adults may encounter difficulty in accurately perceiving non-native speech sounds that use auditory dimensions differently from those to which optimal perception is geared, as in the case of Japanese listeners' discrimination of the English /r-l/ contrast described above. Adults are nevertheless able to learn to discriminate such contrasts more accurately by shifting their relative use of auditory dimensions that are important for distinguishing the two speech sounds [2, 6, 8].

NON-NATIVE SPEECH PERCEPTION

The study of non-native perception typically features listeners who differ with respect to their experience with the non-native language. *Naïve* or *inexperienced (non-native) listeners* are individuals for whom the linguistic variety of the speech signal is unfamiliar, i.e., listeners who have no or very limited experience with it, while *second-language (L2) listeners* or *learners* are individuals who are actively involved in learning a language.

Under the umbrella of non-native speech perception, two related areas of research have emerged, *cross-language* and *second-language (L2) speech perception*. The former generally refers to the processing of the non-native speech signal in terms of one's native language, which has long been held to reveal the auditory dimensions that are (ir)relevant for native listeners as well as changes in infants' early abilities [19]. Studies in the latter area, by contrast, typically investigate adult learners with varying degrees of experience with the L2 [20, 21], including the many factors that can affect learning, e.g., formal language instruction, motivation, length of residence in an L2-speaking country among others [22]. Additionally, cross-language perception with naïve adult listeners can reveal how beginning learners will perceive the sounds of an L2, referred to as the 'initial state' [7, 8].

A further area of non-native speech perception, though perhaps not conventionally grouped within it, is *cross-dialect speech perception*. Below we will see that findings from this area of research are also very relevant to cross-language and L2 speech perception.

Cross-dialect speech perception

The early phonetic ability of infants is apparent in cross-dialect speech perception as infants are able to discriminate between the dialect around them and unfamiliar dialects of the same language [23]. However, younger toddlers find it more difficult to recognise words spoken in an unfamiliar non-native dialect than older toddlers [24], indicating that adaptation to non-native dialects occurs with phonological development at the onset of word learning. An apparent 'bias' toward listeners' native dialect can extend into adulthood. For instance, listeners within and between English-speaking countries may use acoustic cues, such as the F1, F2 and duration of vowels, in slightly different ways to identify the same phonological categories [6, 25]. This is also the case for Spanish and Portuguese speakers from Latin American and Europe [26].

Listeners are able to adapt to unfamiliar dialects, even after limited exposure [27], though phonetic similarity between listeners' and speakers' dialects facilitates adaptation [28, 29]. However, some sound contrasts may be persistently problematic for non-native dialect listeners, especially when a phonologically equivalent contrast does not exist in listeners' native dialect. [30-35].

One example is the lack of the English vowel /ʌ/ in Northern British English dialects where words such as 'book' and 'buck', which contain the phonetically and phonologically distinct /ʊ/ and /ʌ/ vowels in Southern British English dialects, are realised as [ʊ]. This lack of separation between the two phonological categories is mirrored in speech perception: Northern listeners' 'exemplars' of Southern /ʌ/ (based on duration, F1 and F2 values) are very unlike how Southern speakers produce this vowel and resemble /ʊ/ [36].

Northern listeners who have lived in the South of England for an extended period of time are able to shift their exemplar locations so that Southern /ʌ/ exhibits higher F1 frequencies than /ʊ/, indicating a phonologically distinct vowel. Nevertheless, experienced Northern listeners' exemplar locations do not accurately match those of Southern listeners or how the vowel is produced by Southern speakers, meaning Southern /ʌ/ is still problematic for even experienced Northern listeners.

Cross-language speech perception

Recent research in cross-language speech perception has also examined perception of sounds in different varieties of an unfamiliar language. Escudero and Chládková [37] show that Spanish listeners are sensitive to differences in F1 and F2 values of American English and Southern British English vowels, e.g., Southern British English /æ/ is perceived to be more similar to Spanish /a/, whereas American English /æ/, which exhibits lower F1 values than in Southern British English, is perceived to be more like Spanish /e/. This suggests different initial states for Spanish learners of the two English dialects.

Listeners whose native vowel system is much smaller than that of the non-native language are more likely to perceive some non-native vowels as instances of the same native category, often leading to poor discrimination accuracy. However, this also depends on the specific acoustic properties of native vowels. For example, Salento Italian and Peruvian Spanish both exhibit the same five-vowel system, but listeners perceive some Standard Southern British English vowels differently. Escudero *et al*. [38] show that Standard Southern British English /ɒ-ɔː/ are mapped onto a single category /o/ by Salento Italian listeners but onto both /o/ and /u/ by Peruvian Spanish listeners. This suggests greater discrimination accuracy than Salento Italian listeners and different initial states for both groups of listeners [38].

Different initial states may be observed due to listeners' different native dialects. For example, a major difference between the Bohemian and Moravian Czech dialects is the /ɪ-iː/ contrast: Bohemian Czech /iː/ has a lower F1 and is longer than /ɪ/, while only durational differences contrast these two vowels in Moravian Czech. Chládková and Podlipský [39] show that naïve Bohemian Czech listeners perceive the Dutch vowels /i/ and /ɪ/ to be most similar to their Czech /ɪ/ and /iː/ categories, respectively, while Moravian Czech listeners perceive the two Dutch vowels to be most similar mainly to their Czech /ɪ/ category. It is predicted therefore that Moravian Czech listeners' discrimination of the Dutch contrast will be poorer and therefore more difficult for them to learn than for Moravian Czech individuals.

Second-language speech perception

Predictions based on the initial states from cross-language research appear to be borne out in second-language learners.

Firstly, L2 learners' perceptual development depends on how speech sounds are contrasted in their L2 environment, as proposed in the LP model. In speech production, the English /iː-ɪ/ contrast is realised by F1 differences in Scottish English and by F1 and durational differences in Southern British English. Escudero and Boersma [6] presented Spanish learners of English with synthetic stimuli that varied in equal auditory steps along duration, F1 and F2 dimensions (covering the ranges of F1 and F2 of naturally produced Scottish English /iː/ and /ɪ/) and instructed learners to select the English vowel they heard by clicking on a picture representing /ɪ/ or /iː/. Spanish learners who were learning English in the South of England made greater use of duration to perceive the contrast, while those learning English in Scotland tended to use F1 and F2. While Spanish learners of Scottish English used auditory dimensions also relevant for perceiving vowels in Spanish, those learning Southern English made use of a new auditory dimension not used in Spanish, namely duration, demonstrating different learning strategies depending on the dialect being learned.

Secondly, the native dialects of learners may affect how speech sounds and contrasts are learned in a L2. Escudero *et al*. [40] found that Flemish Dutch and North Holland Dutch learners of English exhibit different levels of errors identifying the English vowels /ɛ/ and /æ/ due different confusion patterns arising from different native dialects. Likewise, Escudero and Williams [41] show that native dialect affects Peruvian and Iberian Spanish learners of Dutch on several vowels and contrasts. Indeed, learners' native dialect was generally a better predictor of L2 discrimination accuracy than measures of L2 proficiency. Both of these studies demonstrate that differential early experiences with their native language influenced subsequent L2 learning.

Finally, L2 learners are able to achieve discrimination accuracy comparable to that of native listeners, but the ways in which auditory dimensions are integrated in speech perception may not resemble that of native listeners. For instance, Escudero *et al*. [42] demonstrate that Spanish learners of Dutch can successfully categorise Dutch /aː-ɑ/ tokens as accurately as native Dutch listeners, but their cue-weighting is very different. While both native Dutch listeners and Spanish learners rely on duration, F1 and F2, Spanish learners exploit duration more heavily than spectrum whereas Dutch listeners use spectrum more heavily than duration. Nevertheless, Spanish listeners' category boundary of Dutch /aː-ɑ/ is less clearly defined, suggesting some uncertainty.

Furthermore, cue-weighting in a manner similar to Dutch listeners does not guarantee accurate categorization as Escudero *et al*. [42] report on a group of naïve German listeners who also performed the same task. Like Dutch listeners, they weighted spectrum more heavily than duration, which suggests German individuals will learn the Dutch contrast in a different way from Spanish learners. However, German listeners were less accurate at categorizing Dutch /aː-ɑ/ tokens than native Dutch listeners, suggesting experience with the Dutch language is of course necessary for more accurate discrimination.

CONCLUSIONS

The present review has highlighted the development of native speech perception from infancy into adulthood and its influences on non-native speech perception. Escudero's [7, 8] Linguistic Perception model formalises this development as a perception grammar in which individuals map relevant auditory dimensions onto speech sounds in accordance with the acoustic dimensions used in speech production in their speech environments (referred to as 'optimal perception'). Individuals' differing experiences with language, including varieties of the same language, thus influence which auditory dimensions are used and how these are used in the perception of speech sounds in unfamiliar non-native languages, different dialects and second languages.

REFERENCES

[1] P. Iverson, P. Kuhl, R. Akahane-Yamada, E. Diesch, Y. Tohkura, A. Ketterman and C. Siebert, "A perceptual interference account of acquisition difficulties for non-native phonemes", *Cognition* **87**, B47-B57 (2003)

[2] P. Iverson, V. Hazan and K. Bannister, "Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults", *Journal of the Acoustical Society of America* **118**, 3267-3278 (2005)

[3] P. K. Kuhl, B. T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola and T. Nelson, "Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e)", *Philosophical Transactions of the Royal Society B* **363**, 979-1000 (2008)

[4] C. T. Best, "A direct realist view of cross-language speech perception", in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, York Press, Timonium MD, 1995, pp. 171-204

[5] J. Flege, "Second language speech learning: Theory, findings and problems", in W. Strange (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* York Press, Timonium MD, 1995, pp. 229-273

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

[7] P. Escudero, *Linguistic perception and second-language acquisition: Explaining the attainment of optimal phonological categorization*, LOT Dissertation Series 113, Utrecht University 2005.

[8] P. Escudero, "The linguistic perception of SIMILAR L2 sounds", in P. Boersma and S. Hamann (eds.), *Phonology in Perception*, Mouton de Gruyter, Berlin, 2009, pp. 152-190

[9] P. Escudero, "Phoneme discrimination", in P. J. Brooks and V. Kempe (eds.), *Encyclopedia of Language Development*, Sage, Thousand Oaks CA, 2014, pp. 444-447

[10] H. Goto, "Auditory perception by normal Japanese adults of the sounds 'L' and 'R'", *Neuropsychologia* **9**, 317-323 (1971)

[11] J. Werker and R. Tees, "Phonemic and phonetic factors in adult cross-language speech perception", *Journal of the Acoustical Society of America* **75**, 1866-1878 (1984)

[12] C. Best and G. McRoberts, "Infant Perception of non-native contrasts that adults assimilate in different ways", *Language and Speech* **46**, 183-216 (2003)

[13] P. Eimas, "Auditory and phonetic coding of the cues for speech: discrimination of the /r-l/ distinction by young infants", *Perception and Psychophysics* **18**, 341-347 (1975)

[14] A. Liberman and I. Mattingly, "The motor theory of speech perception revised", *Cognition* **21**, 1-36 (1985)

[15] P. Kuhl, "Discrimination of speech by nonhuman animals: basic auditory sensitivities conductive to the perception of speech-sound categories", *Journal of the Acoustical Society of America* **70**, 340-349 (1981)

[16] P. Jusczyk, D. Pisoni, A. Walley and J. Murray, "Discrimination of relative onset time of two-component tones by infants", *Journal of the Acoustical Society of America* **67**, 262-270 (1980)

[17] P. Boersma, P. Escudero and R. Hayes, "Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories", *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 3-9 August 2003, pp. 1013-1016.

[18] J. Maye, J. F. Werker and L. A. Gerken, "Infant sensitivity to distributional information can affect phonetic discrimination", *Cognition* **82**, B101-B111 (2002)

[19] J. Werker and C. Lalonde, "Cross-language speech perception: initial capabilities and developmental change", *Developmental Psychology* **24**, 672-683 (1988).

[20] I. MacKay, J. Flege, T. Piske and C. Schirru, "Category restructuring during second-language speech acquisition", *Journal of the Acoustical Society of America* **110**, 516-528 (2001)

[21] J. Flege and I. MacKay, "Perceiving vowels in a second language", *Studies in Second Language Acquisition* **26**, 1-34 (2004)

[22] C. Best and M. Tyler, "Nonnative and second-language speech perception: Commonalities and complementaries", in O.-S. Bohn and M. Munro (eds.), *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*, John Benjamins, Amsterdam, 2007, pp. 13-24

[23] J. Butler, C. Floccia and G. Goslin, "Infants' discrimination of familiar and unfamiliar accents in speech", *Infancy* **16**, 392-417 (2011)

[24] C. Best, M. Tyler, T. Gooding, C. Orlando and C. Quann, "Development of phonological constancy: toddlers' perception of native and Jamaican-accented words", *Psychological Science* **20**, 539- 542 (2009)

[25] A. Ghonim, J. Lim, J. Smith and J. Wolfe, "The division of the perceptual vowel plane for different accents of English and the characteristic separation required to distinguish vowels", *Acoustics Australia* **41**, 160-164 (2013)

[26] K. Chládková and P. Escudero, "Comparing vowel perception and production in Spanish and Portuguese: European versus Latin American dialects", *Journal of the Acoustical Society of America* **131**, EL119-EL125 (2012)

[27] J. Maye, R. N. Aslin and M. K. Tanenhaus, "The weckud wetch of the wast: lexical adaptation to a novel accent", *Cognitive Science* **32**, 543-562 (2008)

[28] M. Sumner and A. Samuel, "The effect of experience on the perception and representation of dialect variants", *Journal of Memory and Language* **60**, 487-501 (2009)

[29] J. Le, C. Best, M. Tyler and C. Kroos, "Effects of non-native dialects on spoken word recognition", *Proceedings of Interspeech 2007*, Antwerp, Belgium, 27-31 August 2007, pp. 1589-1592

[30] C. Clopper, "Effects of naturally-occurring segmental ambiguity on cross-modal priming of minimal pairs", *Journal of the Acoustical Society of America* **129**, 2659 (2011)

[31] S. Dufour, N. Nguyen and U. H. Frauenfelder, "The perception of phonemic contrasts in a non-native dialect", *Journal of the Acoustical Society of America* **121**, EL131-EL136 (2007)

[32] S. Dufour, N. Nguyen and U. H. Frauenfelder, "Does training on a phonemic contrast absent in the listener's dialect influence word recognition?", *Journal of the Acoustical Society of America* **128**, EL43-EL48 (2010)

[33] A. Brunellière, S. Dufour, N. Nguyen and U. H. Frauenfelder, "Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception", *Cognition* **111**, 390-396 (2009)

[34] A. Brunellière, S. Dufour and N. Nguyen, "Regional differences in listener's phonemic inventory affect semantic processing: A mismatch negativity (MMN) study", *Brain and Language* **117**, 45-51 (2011)

[35] B. Conrey, G. F. Potts and N. A. Niedzielski, "Effects of dialect on merger perception: ERP and behavioural correlates", *Brain and Language* **95**, 435-449 (2005)

[36] B. Evans and P. Iverson, "Vowel normalization for accent: an investigation of best exemplar locations in northern and southern British English sentences", *Journal of the Acoustical Society of America* **115**, 352-361 (2004)

[37] P. Escudero and K. Chládková, "Spanish listeners' perception of American and Southern British English vowels: Different initial stages for L2 development", *Journal of the Acoustical Society of America* **128**, EL254-EL260 (2010)

[38] P. Escudero, B. Sisinni and M. Grimaldi, “The effect of vowel inventory and acoustic properties in Salento learners of Southern British English vowels”, *Journal of the Acoustical Society of America* **135**, 1577-1584 (2014)

[39] K. Chládková and V. Podlipský, “Native dialect matters: Perceptual assimilation of Dutch vowels by Czech listeners”, *Journal of the Acoustical Society of America* **130**, EL186-EL192 (2011)

[40] P. Escudero, E. Simon and H. Mitterer, “The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception”, *Journal of Phonetics* **40**, 280-288 (2012)

[41] P. Escudero and D. Williams, “Native dialect influences second language vowel perception: Peruvian versus Iberian learners of Dutch”, *Journal of the Acoustical Society of America* **131**, EL406-EL412 (2012)

[42] P. Escudero, T. Benders and S. Lipski “Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners”, *Journal of Phonetics* **37**, 452-465 (2009)