

# Methodological issues in assessing L2 perceptual phonological competence

Joan C. Mora, University of Barcelona

**1 Introduction** Current L2 speech learning models such as the Speech Learning Model (SLM; Flege 1995) or the Perceptual Assimilation Model (PAM; Best 1995) aim at accounting for learners' differences in their ability to perceive L2 contrastive vowels and consonants on the basis of first (L1) and second (L2) language experience and the degree of perceived phonetic similarity between L1 and L2 speech sounds. The SLM, for example, hypothesises that when phonetic differences between L2 and L1 phonetically similar sounds are not discerned, a perceptual link between these sounds is created through equivalence classification, thus blocking the formation of new phonetic categories for L2 sounds, which inevitably leads to lack of pronunciation accuracy and foreign accent in L2 speech. Thus, the SLM adheres to the notion that *perception leads production* in L2 speech learning and assumes that segments are produced "only as accurately as they are perceived" (Flege 2003: 344). An obvious implication of this approach for foreign language (FL) pronunciation instruction is that perceptual training should constitute an essential component in the design of pronunciation tasks, but it also has important implications for L2 phonological competence assessment in that perceptual competence may be assumed to be an indicator of the state of development of the learner's L2 phonological system. However, assessing the perceptual phonological competence of L2 learners is methodologically problematic because learners' performance is extremely sensitive to task variables. Investigating the effect of task design variables on FL learners' perceptual ability may prove helpful in identifying the test formats that can best measure FL learners' phonological competence.

This paper examines several methodological issues in the design of testing instruments to assess perceptual phonological competence in the light of findings of previous and current research on the ability of advanced Catalan/Spanish late learners of English to discriminate L2 phonemic contrasts varying in degree of perceptual difficulty. The performance of two groups of learners on two categorial discrimination tasks differing in speech stimuli type and presentation format were compared. Contrary to our expectations and previous research on the effect of task design on categorial discrimination (Højen & Flege 2006; Mora 2005), an AXB task consisting of non-words elicited in a carrier sentence yielded higher discrimination scores than an AX task consisting of words elicited in isolation. These findings suggest that the use of non-words in the AXB task made the target vowel and consonant contrasts easier to perceive because it changed the learners' attentional focus to an acoustic rather than phonemic level of perception, and this enhanced their ability to perceive differences between L2 sounds that belong to different L2 sound categories.

**2 Assessing perceptual phonological competence in a FL instructional setting** The amount of input L2 learners receive in a FL instructional setting is not only scarce and often foreign accented, opportunities for perceiving authentic L2 speech sounds in a wide variety of contexts is also severely constrained by the use of a very limited range of lexical diversity. These learning context variables should be considered when assessing L2 learners' perceptual competence by adjusting the sensitivity of perceptual tasks. The type of stimuli used and the way they are presented in identification and discrimination

tasks are likely to pose different perceptual difficulties for learners in immersion and formal instruction settings.

In cross-language speech perception research L2 learners' phonological competence is normally assessed through identification and discrimination tasks (Beddor & Gottfried 1995; Hamsberger 2001). Identification tasks, involving the assignment of a category label to a given sound stimulus, are useful in providing information about the interaction between the learner's L1 and L2 phonetic systems and may be used to predict the kind of within-category and between-category discrimination difficulties that prevent FL learners from attaining native-like perceptual (and productive) phonological competence. Whereas identification tasks are used to predict degree of perceptual difficulty, the assessment of L2 learners' perceptual phonological competence is normally carried out through AX or AXB categorial discrimination tasks that test learners' ability to discriminate between contrastive L2 vowels or consonants. From a SLM perspective, learners' relative degree of success in discriminating contrastive L2 sounds reflects their overall level of phonological competence because accuracy in production is dependent on learners' ability to discern differences between L2 sounds. Failure to discriminate between contrastive L2 speech sounds blocks the formation of new phonetic categories for those sounds (they are assimilated to L1 phonetic categories instead), which results in lack of native-like accuracy in L2 speech sound production. Although perceptual difficulty discriminating L2 sounds is largely predictable from the degree of phonetic similarity between L2 and L1 sounds (Best 1995; Flege 1995), which can be established directly through acoustic measurements or indirectly through category goodness judgements, variation in learners' performance on discrimination tasks cannot be fully accounted for by degree of phonetic (dis)similarity or the relative difficulty of the type of segmental contrasts they are asked to discriminate; the type of stimuli used and the way they are presented also determine to a great extent learners' performance on categorial discrimination tasks. At least 6 task design variables have been found to affect learners' performance (Beddor & Gottfried 1995; Cowan & Morse 1986; MacMillan et al. 1988; Schouten & Hensen 1992): (1) task type: AX, ABX, AXB, 4IAX; (2) using words vs. non-words; (3) using words/non-words vs. isolated sound segments (or syllables); (4) using natural vs. modified or synthetic stimuli; (5) fully randomised vs. blocked stimuli presentation; (6) long (1000-2000 ms.) vs. short (0-750 ms) inter-stimulus intervals (ISI). These task design variables may be adjusted to match the needs of the research questions being asked. For example, a long ISI is typically used in between-category discrimination tasks because it forces hearers to rely on pre-existing mental representations of sounds, whereas a short-enough ISI would allow hearers to compare the stimuli in auditory sensory memory. These variables can also be systematically modified to make them phonetically sensitive to listeners differing in L1 background and L2 proficiency, or to avoid ceiling effects (Højen & Flege 2006).

### **3 Effects of task type and lexical knowledge on learners' discrimination ability**

Whereas in immersion settings learners' performance typically reflect their ultimate level of attainment in L2 speech learning, in the FL learning context learners' perceptual abilities are in a state of constant development and the focus of assessment is on capturing significant gains in competence over time. In order to be able to measure improvement and observe gains in perceptual ability over time, discrimination tasks need to be carefully designed to avoid obtaining scores that are either at ceiling or at chance level. One solution to this methodological problem is selecting L2 sound contrasts on the basis of their predicted degree of discrimination difficulty so that "easy-" and "difficult-to-discriminate" contrasts are avoided. However, L2 learners have been reported to obtain near-ceiling scores in AXB discrimination tasks even for "difficult" L2 sound contrasts (i.e. those assimilated to a single L1 category). Another possible solution is to develop a

convenient variant of the discrimination task by modifying some of its parameters. Højen & Flege (2006), for example, in a study of L2 vowel discrimination, modified the inter-stimulus interval duration (0 vs. 1000 ms), the block structure (small vs. large stimulus range), and introduced *F0* variation in physically different stimuli to lower discrimination scores. In this paper we report on the results of three perception studies measuring the perceptual ability of three similar groups of bilingual Catalan/Spanish first-year university students (advanced EFL learners) through three discrimination tasks with different formats and the effect of task variables is discussed.

The first study investigated learners' differential gains in linguistic competence as a result of two different learning contexts: a period of formal instruction that included no pronunciation training (FI) in their home university in Barcelona and a three-month stay abroad (SA) term in an English-speaking country. Perceptual accuracy was measured upon students' enrolment at university (T1), after two terms of FI (T2) and after a three-month SA term in an English-speaking country (T3). The participants in this study (Mora 2007) showed considerable differences in discrimination ability across contrasts, ranging from near-chance scores for word-final /s-z/ to almost native-like performance for /æ-ɑ:/ (see Table 1). The discrimination task used was a forced-choice AX test consisting of 144 word-pairs (108 minimal pairs + 36 distractor same-word pairs) presented in 2 fully randomized blocks (1 block for vowel contrasts and 1 for consonant contrasts) at 1 second ISIs and inter-trial intervals (ITI) of 2 seconds. All 12 word-pairs used per phonemic contrast were different. At T2, percent correct scores for 5 out of 8 phonemic contrasts was over 80% (3 of them over 85%) and at T3 slight non-significant gains in discrimination ability were observed. These results would seem to suggest that the FI period had a greater effect on the subjects' ability to discriminate English phonemic contrasts than the SA period. However, it may also be the case that the design of the discrimination test produced scores that were too high at T2 to be able to capture perceptual gains at T3, i.e. a test instrument yielding lower percent correct discrimination scores at T2 might have been more successful in capturing perceptual gains at T3.

Mean %	/i:-ɪ/	/æ-ʌ/	/æ-ɑ:/	/ɪ-ə/	/e-eə/	/t-d/	/s-z/	/tʃ-dʒ/	/d-ð/	
N=25	T1	72.67 (20.35)	77.33 (26.19)	94 (10.63)	92.33 (9.6)	60.33 (15.27)	72.67 (20.06)	48.33 (23.56)	81.67 (10.49)	57.33 (17.23)
	T2	81.67 (21.25)	83 (20.76)	93.67 (13.67)	96.33 (6.4)	70 (12.27)	79.67 (15.61)	57.33 (25.95)	85.33 (11.1)	66.67 (15.77)
	T3	81.67 (23.81)	83.67 (23.75)	95 (12.5)	95.67 (7.65)	76 (11.86)	83.67 (18.4)	59.67 (25.65)	87.33 (11.81)	69.67 (19.97)

Table 1. Mean % correct discrimination (SD in parenthesis) in a AX task (from Mora 2007).

In this discrimination task participants' knowledge of the lexical items in the test was not controlled. Previous research has shown that listeners make use of lexical knowledge in the phonological categorization of sounds (e.g. Norris et al. 2003), suggesting that the use of non-words might have lowered the discrimination scores in the study reported above. In order to investigate this possibility, a parallel second study was carried out with a similar population (74 first-year university students) using exactly the same discrimination test format, but substituting half of the words and distractors by corresponding non-words. The results of this study (Mora 2005) revealed an overall effect of lexical knowledge: learners were more successful in perceiving nonnative contrasts in minimal pairs with known words than in minimal pairs with non-words (see Table 2). However, a closer look at the distribution of scores shows that the lexical knowledge advantage does not hold for all contrasts, with only 5 of the 9 phonemic contrasts examined being better perceived in words than in non-words. Interestingly,

significant differences in discrimination scores between the word and the non-word conditions occur irrespective of the large differences in percent correct discrimination across phonemic contrasts. This indicates that the lexical knowledge advantage is fairly robust across phonemic contrasts and suggests that the use of non-words may be a useful strategy to make discrimination tasks more sensitive for advanced L2 learners.

Mean %		/i:-ɪ/	/æ-ʌ/	/æ-ɑ:/	/ɪ-ə/	/e-eə/	/t-d/	/s-z/	/tʃ-dʒ/	/d-d/
N=74	Wds	83.33 (17.88)	88.96 (13.31)	96.40 (11.46)	93.24 (12.92)	67.79 (18.16)	88.51 (14.31)	78.60 (23.65)	70.27 (18.35)	48.65 (22.54)
	NWds	84.68 (20.23)	72.75 (21.11)	96.84 (8.13)	86.71 (14.06)	82.43 (10.99)	82.88 (15.10)	61.71 (27.82)	47.30 (19.71)	53.15 (16.93)

Table 2. Mean % correct discrimination in a AX task (adapted from Mora 2005).

Another study, with the same longitudinal research design as the first study reported above (3 data collection times before and after FI and SA periods) and an equivalent group of participants (first-year university students, advanced EFL learners) was carried out. The participants were tested on a discrimination task that included 4 of the 9 phonemic contrasts in the two studies reported above, but the following conditions in the discrimination task were changed: (1) the stimuli were English non-words; (2) the stimuli were elicited in a carrier phrase (not pronounced in isolation); and (3) The task was an AXB (not AX) test, 4 orders (ABB, AAB, BAA, BBA) with none of the 6 A or B non-word stimuli being physically identical or appearing in the same position. The scores obtained (Table 3) reveal the same pattern as those obtained in the first study reported above (see Table 1). The results seem to confirm that the two-term FI period produced greater gains in perceptual competence than the SA term.

Mean %		/i:-ɪ/	/æ-ʌ/	/s-z/	/tʃ-dʒ/
N=32	T1	93.01 (6.75)	92.74 (8.74)	72.04 (10.21)	61.69 (9.34)
	T2	97.78 (3.95)	96.87 (6.07)	88.80 (15.21)	83.33 (22.99)
	T3	94.44 (7.03)	93.93 (7.22)	75.00 (9.58)	65.53 (10.93)

Table 3. Mean % correct discrimination in the AXB task with non-words.

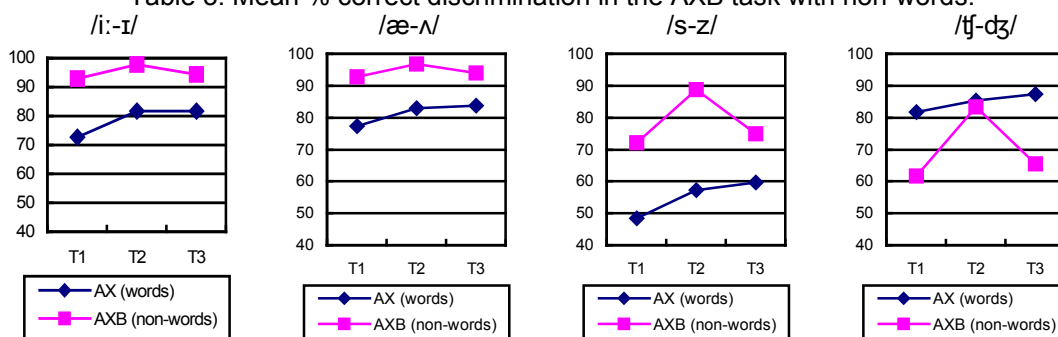


Figure 1. Mean % correct discrimination scores in the AX and AXB tasks.

**4 Conclusion** The findings of the studies reported here suggest that for Catalan/Spanish advanced learners of English the use of an AXB discrimination task (as opposed to an AX task) and non-words (as opposed to words), contrary to our expectations, did not produce the effect of lowering L2 learners' discrimination scores. A tentative explanation is that the use of non-words in the AXB task made the phonemic contrasts easier to

perceive because it changed the learners' attentional focus to an acoustic rather than phonemic level of perception, and this facilitated the perception of acoustic differences between L2 sounds belonging to different L2 sound categories. Further research should examine whether within-category discrimination tasks constitutes a more sensible tool for capturing gains in perceptual ability for advanced learners. Phonetic training in segmental L2 speech perception may also provide us with an interesting context for testing the effect of the task variables discussed in the present paper.

## 5 References

- Beddor, Patrice S. & Gottfried, Terry L. (1995) Methodological Issues in Cross-Language Speech Perception Research with Adults. In Strange, W. (ed) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Timonium, MD: York Press. pp. 207-232.
- Best, Catherine. (1995) A direct realist view of cross-language speech perception. In Strange, W. (ed) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Timonium, MD: York Press. pp. 171-204.
- Cowan, Nelson & Morse, Philip A. (1986) The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*. 79 (2), pp. 500-507.
- Flege, James E. (1995) Second-language speech learning: theory, findings and problems. In Strange, W. (ed) *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Timonium, MD: York Press. pp. 229-273.
- Flege, James, E. (2003) Assessing constraints on second-language segmental production and perception. In Schiller, N. & Meyer, A. (eds) *Phonetics and phonology in Language Comprehension and Production*. pp. 319-355.
- Hamsberger, James D. (2001) On the relationship between identification and discrimination of non-native nasal consonants. *Journal of the Acoustical Society of America*. 110 (1), pp. 489-503.
- Højen, Anders & Flege, James E. (2006) Early learners' discrimination of second-language vowels. *Journal of the Acoustical Society of America*. 119 (5), pp. 3072-3084.
- Macmillan, Neil A., Goldberg, Rina F. & Braid, Louis D. (1988) Resolution for speech sounds: basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America*. 84 (4), pp. 1262-1280.
- Mora, Joan C. (2005) Lexical knowledge effects on the discrimination of non-native phonemic contrasts in words and non-words by Catalan/Spanish bilingual learners of English. *Proceedings of the ISCA Workshop on Plasticity in Speech Perception*. UCL, London, UK. pp. 43-46.
- Mora, Joan C. (2007) Learning context effects on the acquisition of a second language phonology. In Pérez-Vidal, C., Juan-Garau, M. & Bel, A. (eds) *A Portrait of the Young in the New Multilingual Spain*. Clevedon, UK: Multilingual Matters. pp. 241-263.
- Schouten, M. E. H. & Hesse, A. J. van (1992) Modeling phoneme perception. I: Categorical perception. *Journal of the Acoustical Society of America*. 92 (4), pp. 1841-1855.

## Acknowledgements

This research was supported by research grant HUM2004-05442-C02-01/FILO from the Spanish Ministry of Science and Education. I would like to thank Brian Mott for reading aloud the stimuli for the discrimination tests and all the students who kindly volunteered to do the perception tests.