

“Age” effects on second language acquisition

James Emil Flege, Ian R. A. MacKay***

*University of Alabama at Birmingham, **University of Ottawa
flegejames@yahoo.com, imackay@uottawa.ca

ABSTRACT

Four hypotheses regarding “age” effects on ultimate L2 proficiency are considered here. All four are found to have some predictive power, but none is perfect. This suggests that age effects arise from multiple factors that co-vary with age in ways that are not yet fully understood.

Keywords: Age, AOA, foreign accent, critical period, speech.

1. INTRODUCTION

That “earlier is better” for learning a second language (L2) has been demonstrated convincingly from the 1980s on. An “age” factor seems to be especially strong for L2 speech perception and production. Hundreds of studies have examined how particular vowels and consonants (“sounds”, for short) are identified or discriminated by listeners, how L2 learners articulate the sounds making up L2 words, and the effects of inaccurate L2 articulation on native listeners’ attempts to recognize words spoken with foreign accent (FA).

Surprisingly, what the general public has gleaned about L2 speech learning is not too different from the “received wisdom” of academics not specializing in L2 speech research. It is still common to hear it said that everyone who begins to learn an L2 after puberty is destined to forever speak it with a FA, and that anyone lucky enough to begin learning an L2 before the end of a “critical” period will learn the L2 effortlessly, rapidly, and perfectly. This view not without an element of truth, but it is mistaken in important ways. The aim of this talk, then, is to briefly review some research examining age effects on overall success in learning an L2, and to underscore important questions that remain unanswered.

1.1. Age?

In L2 acquisition research, “age” refers to the *chronological* age at which L2 learning began. My own work examined immigrants whose age of L2 learning was their age of arrival (AOA) in the host country. One study, for example, examined native speakers of Italian who learned English in Canada after emigrating there from Italy (Flege et al. 1995). Another examined Koreans who learned English in the United States (Flege et al. 1999). With but few exceptions, contact with the L2 was immediate and plentiful for these research participants, for they were compelled to learn their L2 (English) due to its social and economic dominance in the surrounding community.

In studies such as these, AOA marked participants’ first substantial contact with the L2 as spoken by native speakers of the L2. A few participants who arrived as adolescents or young adults reported having studied English briefly in school before emigrating, but there is no evidence that this previous experience exerted a lasting effect. To ensure that our research focused on sequential rather than simultaneous bilinguals, we excluded individuals who arrived prior to the age of 2 years.

AOA has proven to be a useful research variable for several reasons. It can be reported accurately; participants differing in AOA are fairly easy to recruit; and constructing groups that differ in AOA ensures robust outcomes. The real importance of AOA, however, is its presumed association with “causative” variables thought likely to vary with AOA. These include variation in neural maturation (itself thought to influence degree of neural plasticity), the state of development of native language phonetic categories, and the kind and/or amount of L2 input typically received.

AOA varies in a near-continuous fashion in immigrant populations inasmuch as whole families often emigrate together. Some research studies have adopted a correlation approach, evaluating the effect of AOA in samples of participants selected randomly or on the basis of some other variable. Better controlled

research has used AOA as a selection variable. A canonical research design, for example, would be one contrasting “early” learners having AOAs of, say, 2 to 10 years to a group of “late” learners having AOAs of, say, 15 to 23 years. A more ambitious design was that of Flege et al. (1995), who selected 10 groups of 24 participants each. The groups’ mean AOAs ranged from 3 (*range* 1.9 to 4.1) to 21 years (*range* 20.2 to 23.2). One aim of this study, and a similar study examining Korean immigrants to the U.S. (Flege et al. 1999), was to identify the AOA of the first non-native group to differ significantly from the native English (NE) comparison group.

As far as I know, all research to date has demonstrated strong effects of AOA. Consider, for example, the seminal study of Johnson and Newport (1989). These authors examined knowledge of 12 grammatical structures said to be the most important of English. The stimuli consisted of both grammatical (*Last night the old lady died in her sleep*) and ungrammatical versions (**Last night the old lady die in her sleep*) of 138 English sentences. The participants were 46 Chinese and Korean speakers, all students or professors at an American university who had lived in the U.S. for at least 3 years. Their task was to label each sentence as “grammatical” or “ungrammatical”. A strong correlation between AOA and the percent correct scores was obtained: the later the arrival, the lower were the scores.

A second example is a study by Flege (1991), who measured a well-known acoustic phonetic dimension in speech production, voice-onset time (VOT). This study compared early to late native Spanish learners of English. The early learners were found to produce word-initial English /t/ with mean VOT values that were very similar to English monolinguals’ whereas the late learners’ VOT values were intermediate to the mean values obtained for English and Spanish monolinguals.

2. EXPLAINING “AGE” EFFECTS

Although age (AOA) effects are robust, the best explanation for these effects remains uncertain, even controversial. With this in mind, I will review four general hypotheses culled from the literature that might be used to explain age effects.

2.1. The maturational constrain hypothesis (H1)

H1 states that endogenous changes that occur as humans mature render less effective mechanisms that subserve language acquisition. DeKeyser (2000: 518-519) observed strong effects of AOA on grammaticality judgment test scores. He concluded that such effects arose as the consequence of a “*severe decline [in] ability to induce abstract patterns implicitly*” which is an “*inevitable consequence of fairly general aspects of neurological maturation*”. In a similar vein, Scovel (2000) concluded that age effects have a “*neuromotor*” etiology, attributing the poorer performance by late than early learners to the reduction of cerebral “*plasticity*” after the closing of a “*critical period*” for L2 learning.

One strength of H1 is that it correctly predicts differences between early and late learners as well the strong AOA-L2 performance correlations observed in many studies. H1 also correctly predicts the results obtained in an analysis of data obtained by Flege et al. (1995) and Flege et al. (1999). These studies evaluated overall degree of perceived FA in sentences spoken by Italian and Korean immigrants to Canada and the U.S., respectively. Degree of FA in sentences spoken by the non-native participants were rated along with sentences produced by members of a native English (NE) comparison group. Individual non-native participants who received a mean FA rating that fell within 2.0 *SDs* of the mean rating obtained for the NE group ($n = 24$) were considered to be “accent free” (admittedly a rather lax criterion). As predicted by H1, no Koreans and just one Italian participant with an AOA greater than 12 years were found to be “accent free”.

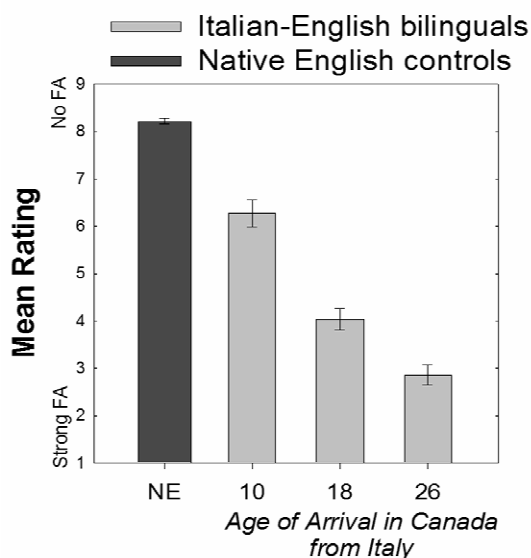
The same analysis, however, revealed a problem for H1. Less than half of the Italians and only about one-quarter of the Koreans with an AOA less than 10 years were found to be accent free. This finding is striking given that all of these participants were educated in English-medium schools, most had been immersed in English for more than 15 years, and nearly all reported using English more than their L1.

The two studies just cited used retrospective developmental designs, that is, examined adults who differed according to when in the past they began learning English. Flege et al. (2005), on the other hand, examined FA in English sentences spoken by groups of Korean children having AOAs that ranged from 5 to 15 years.

After 5 years of immersion in English, the Korean children's sentences still received significantly lower ratings than sentences produced by NE children. The same held true when the same test was performed a year later. A closer look revealed the presence of detectable FAs even in sentences spoken by Korean children who had begun learning English at the age of 6 years and had been immersed in English for more than 5 years.

Another problem for H1 is its generation of a prediction that has been falsified in several studies. If AOA effects are due to the passing of a critical period, one would expect L2 performance to decline as participants' AOAs near the end of a critical period. However, performance should remain stable thereafter because everyone *beyond* the critical period should show equally the ill effects of having passed the critical period. A re-analysis of FA ratings obtained for Italian and Korean immigrants showed, contrary to this expectation, that FAs continue to grow stronger beyond the critical period. For each study (Flege et al. 1995, 1999), AOA-FA correlations were computed for the entire sample (i.e., the 240 participants having AOAs ranging from 2 to 23 years), just for participants having AOAs ranging from 12 to 23, and finally just those having AOAs of 15 to 23 years. As the number of participants decreased, the correlations became weaker but remained significant ($p < .01$) even in the last sample which included only "post critical period" learners.

Figure 1: Ratings of sentences spoken by native English (NE) speakers and groups of Italian immigrants. Each mean value is based on 2,916 ratings (18 participants x 162). The error bars bracket +/- 1 SE.



Flege and MacKay (unpublished) evaluated degree of FA in English samples produced by NE speakers and native Italian groups differing in AOA. As seen in Fig. 1, the pair-wise differences between all four groups were significant ($p < .05$ by Tukey test). H1 might be used to explain the stronger FA of immigrants who arrived in Canada at the age of 18 than 10 years, but not the stronger FA of immigrants who began learning English at 26 than 18 years of age. This is because members of both of the last two groups mentioned had begun learning English after the end of the supposed critical period.

A re-consideration of the Johnson and Newport (1989) morphosyntax data also calls H1 into question. The strongest support obtained by these authors for H1 was a significant correlation between AOA and morphosyntax scores for participants having an AOA of less than 15 years, but not for those having AOAs of 16+. Crucially, the existence of a critical period at age 15 was assumed, not demonstrated. Hakuta and Bialystok (1994) showed that when the "cut point" was arbitrarily shifted to an AOA of 20 years, a moderately strong ($r = -0.49$) correlation between AOA and the morphosyntax scores was obtained for "post-critical period" learners.

Even more strikingly, the results of Flege et al. (1999) suggested that the effect of AOA on grammaticality judgment test scores may disappear when factors confounded with AOA have been controlled. These authors used a test modelled on that of Johnson and Newport (1989). As in the original

study, the test scores showed a strong correlation with AOA. But, as is usually the case in L2 research, AOA was correlated with other variables that might well have influenced the test scores: years of education in English-medium schools, length of residence in the U.S., and self-reported use of English.

Flege et al. (1999) used the matched-subgroup technique in an attempt to unconfound these other potentially important variables. Two groups of 20 participants having mean AOAs of 10 and 17 years were selected from the original sample of 240 without considering any other factor. Not surprisingly, the early learners obtained much higher scores than the late learners did. Two additional groups having the same mean AOA values (viz., 10 and 17 years) were then selected, this time taking care to match the participants in each AOA group for years of education in the U.S. (*mean* 11 years for both). The matching process also eliminated significant between-group differences in years of U.S. residence and self-reported English use. Importantly, the morphosyntax scores for these *matched* groups of early and late learners did not differ (and, in fact, did not even approach significance). This finding leads one to wonder how many “age” effects reported in the literature were really due to factors confounded with AOA and not actually to the age at which L2 learning began.

2.2. The cognitive development hypothesis (H2)

H2 states that L2 learning becomes gradually less effective across the entire life span because cognitive abilities needed for speech and language learning diminish slowly across the life span. H2 is derived from a study by Hakuta et al. (2003). These authors developed estimates of English proficiency based on several responses to the U.S. Census by large numbers of Chinese- and Spanish-speaking immigrants. For both groups, the English proficiency estimates decreased continuously from 10 to 60 years, not just in the second decade of life.

H2 correctly predicts the FA difference in Fig. 1 between immigrants having AOAs of 18 and 26 years. But for H2 to really explain this finding, it will be necessary to identify a specific cognitive change that might be responsible. A more serious limitation for H2, however, is the unlikelihood of its being able to explain the difference between NE speakers and immigrants with an AOA of 10 years (Fig. 1). It is *a priori* difficult to imagine 10-year-olds having undergone a cognitive change that could reduce their ability to learn a second language in a period of life in which the L1 is still developing.

2.3. Changes in L1-L2 interactions (H3)

H3 states that as the L1 system develops, the effect of cross-language phonetic interference becomes stronger because of differences in how the L1 and L2 sound systems interact. This hypothesis is derived from Flege’s Speech Learning Model (e.g., Flege 2002, 2003, 2007, 2009). H3 assumes that L1 categories develop slowly, at least into early adolescence (as shown, for example, by studies examining the recognition of L1 words in noise). According to H3, as L1 categories develop they become stronger “attractors” for sounds later encountered in an L2. This has the effect of making learners less likely to create new categories for L2 sounds that are similar but non-identical to corresponding L1 sounds, even when the cross-language phonetic differences are audible.

H3 generates predictions that can be readily tested. It predicts, for example, that as native Spanish speakers get older, their ratings of the perceived phonetic dissimilarity between tokens of English [th] and Spanish [t] (all realizations of a /t/ phoneme) will decrease. H3 also predicts that the frequency of detections of English-like VOT values inserted into Spanish speech samples will decrease as Spanish monolinguals grow older (and, possibly, that the latencies of correct detections will increase). Most importantly, H3 predicts that as an indirect consequence of the development of L1 phonetic categories, Spanish early learners will be more likely than Spanish late learners to establish a new phonetic category for English /t/, and so will be more likely than late learners to produce English /t/ correctly.

A less obvious prediction—also derived from the SLM—regards what is expected to happen when learners (presumably, mostly late learners) fail to establish a new phonetic category for L2 sounds that differ auditorily from the closest sound in L1. Here the expectation is the development of a composite (merged) category sharing properties of the perceptually “equated” L1 and L2 sounds. This leads to the prediction, for

example, that late Spanish learners of English will tend to produce English /t/ with VOT values that are too short for English, and Spanish /t/ with VOT values that are too long for Spanish.

As already mentioned, Flege (1991) found that early Spanish learners of English produced English /t/ with native-like values whereas late learners produced values intermediate to those observed for Spanish and English monolingual. H3 generates the same predictions for French as for Spanish. Flege (1987) observed “compromise” VOT values for late French-English bilinguals. Also, both French learners of English and American learners of French were found to produce L1 /t/ with values that had moved in the direction of typical L2 values.

A problem for H3 is that little research has been conducted to determine if the perceived dissimilarity of pairs of L1 and L2 sounds actually does decrease as L1 categories develop, and whether such changes—if they occur—predict accuracy of L2 segmental production and/or perception (but see Baker et al. 2002).

2.4. More/better input for early than late learners (H4)

There is evidence that early learners typically get more, and perhaps more authentic, L2 input than late learners do. H4 states that L2 proficiency increases as a function of amount of L2 input. Given the assumption that input varies as a function of language use, it predicts that participants who have used the L2 often over the years will outperform those who have used L2 less often.

The prediction of H4 was supported by research examining Italian immigrants in Canada (Flege and MacKay 2004; Piske et al. 2001, 2002; MacKay et al. 2001). Groups differing in AOA (*means* 8 vs. 20 years) were subdivided according to self-reported frequency of Italian use (*means* of 7% and 10% vs. 43% and 53%). For both early and late learners, a frequent use of Italian (and thus an infrequent use of English) was associated with a significantly poorer identification of word-final English consonants, a poorer discrimination of English vowels, less accurate production of English vowels, and relatively stronger FAs.

Table 1: Characteristics (*means, SDs*) of three groups of Italian immigrants tested by Flege and MacKay (unpubl.). All differences between AOA-10 and AOA-18, but no differences between AOA-18 and AOA-26, were significant ($p < .05$).

	AOA-10	AOA-18	AOA-26
Years of residence in Canada	43(5)	37(7)	33(8)
Years of education in Canada	9(4)	<1(1)	<1(1)
Self-reported % use of English	71(14)	53(15)	47(19)
Self-reported % use of Italian	27(14)	47(15)	52(20)
Self-rated (7-pt) ability to pronounce Italian	5.3(1.1)	6.4(0.8)	6.8(0.5)
Self-rated ability to speak-understand Italian	5.5(1.0)	6.4(0.8)	6.8(0.3)

These findings discredit claims (e.g., DeKeyser 2000: 519) that input differences do not contribute to age effects on L2 speech learning. However, it is not clear at present to what *extent* input differences contribute to age effects inasmuch as L2 research has relied on self-reports of language use rather than taking actual measurements of input. Another problem for H4 is that language use differences do not predict all findings currently available. For example, H4 might explain the difference in Fig. 1 between participants having AOAs of 10 and 18 years but not the difference between those having AOAs of 18 and 26 years. As can be seen in Table 1, the former but not the later pair of groups was likely to have differed in terms of amount of L2 input.

3. CONCLUSIONS

All four hypotheses considered here seemed to have some predictive value, but none offered a convincing explanation of age effect and none was able to account for all of the evidence considered. The possibility therefore exists that all four hypotheses might help explain, to varying degrees, age-related effects seen in the literature.

Future research should be designed so that confounded factors can be eliminated *a priori* or, at the very least, be controlled statistically *a posteriori*. It will be imperative to evaluate variables thought to affect ultimate L2 proficiency directly rather than, as in the past, to manipulate “age” of L2 learning via AOA. Potentially “causative” variables worth investigating include: age-related changes in the perceived relation between sounds in the L1 and L2; variations in the kind of input received (foreign accented?), especially in early stages of L2 learning; and the amount (measured, not estimated!) of L2 input received. Until causal variables have been examined directly and variables confounded with them have been controlled, we can only speculate about the true basis (bases) for age-related differences in ultimate L2 proficiency.

4. REFERENCES

- Baker, W., Trofimovich, P., Mack, M., Flege, J. 2002. The effect of perceived phonetic similarity on non-native sound learning by children and adults. In: Skarabela, B. et al. (eds.). 36-47.
- Bialystok, E., Hakuta, K. 1994. *In other words: The science and psychology of second language acquisition*. New York: HarperCollins.
- Birdsong, D. (ed.) 1999. *Second language acquisition and the critical period hypothesis*. Hillsdale: Lawrence Erlbaum.
- Burmeister, P. et al. (eds.). 2002. *An integrated view of language development: Papers in honor of Henning Wode*. Trier: Wissenschaftlicher.
- Cole, J., Hualde, J. (eds.) 2007. *Laboratory phonology 9*. Berlin: Mouton de Gruyter.
- DeKeyser, R. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22. 499-534.
- Flege, J. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15. 47-65
- Flege, J. 1991. Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America* 89. 395-411.
- Flege, J. 1999. Age of learning and second-language speech. In: Birdsong, D. (ed.). 101-132.
- Flege, J. 2002. Interactions between the native and second-language phonetic systems. In: Burmeister, P. et al. (eds.). 217-244.
- Flege, J. 2003. Assessing constraints on second-language segmental production and perception. In Meyer, N. Schiller (eds.). 319-355.
- Flege, J. 2007. Language contact in bilingualism: phonetic system interactions. In: Cole, J., Hualde, J. (eds.). 353-382.
- Flege, J. 2009. Give input a chance! In: Piske, T., Young-Scholten, M. (eds.). 175-190.
- Flege, J., Birdsong, D., Bialystok, E., Mack, M., Sung, H., Tsukada, K. 2005. Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics* 33. 263-290.
- Flege, J., MacKay, I. 2004. Perceiving vowels in a second language. *Studies in second language acquisition* 26. 1-34.
- Flege, J., Munro, M., MacKay, I. 1995. Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97. 3125-3134.
- Flege, J., Yeni-Komshian, G., Liu, S. 1999. Age constraints on second language learning. *Journal of Memory and Language* 41. 78-104.
- Hakuta, K.; Bialystok, E., Wiley, E. 2003. Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Sciences* 14. 31-38.
- Johnson, J., Newport, E. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21. 60-99.
- MacKay, I., Meador, D., Flege, J. 2001. The identification of English consonants by native speakers of Italian. *Phonetica* 58, 103-125.
- Meyer, A., Schiller, N. (eds.) 2003. *Phonetics and phonology in language comprehension and production*. Berlin: Mouton.
- Piske, T., Flege, J., MacKay, I., Meador, D. 2002. The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica* 59. 49-71.
- Piske, T., MacKay, I., Flege, J. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29, 191-215.
- Piske, T., Young-Scholten, M. (eds.) 2009. *Input matters in SLA*. Bristol: Multilingual Matters.
- Scovel, T. 2000. A critical review of critical period research. *Annual Review of Applied Linguistics* 20. 213-223.
- Skarabela, B., Fish, S., Do, A. (eds.). 2002. *Proceedings of the 26th annual Boston University conference on language development*, Somerville: Cascadilla Press.