STABILITY OF INDIVIDUAL PATTERNS IN LEARNING A SECOND LANGUAGE VOICING CONTRAST

Amy E. Hutchinson, Olga Dmitrieva

Purdue University
hutchi25@purdue.edu, odmitrie@purdue.edu

ABSTRACT

This study examines individual patterns in the production of French voicing categories by learners whose first language (L1) is American English. The focus of the study is on stability of individual production patterns across the first and second language (L2). Twenty-three intermediate-level learners of French were recorded reading an English and a French word list, each with eight minimal pairs contrasting in initial bilabial stop voicing. Participants' accuracy of French production differed greatly, and a relatively successful realization of one member of the contrast (e.g. /b/) did not predict greater success with another (/p/). However, participants who were more likely to prevoice their English /b/s were also more successful at producing prevoicing in French. In English, participants consistently relied on VOT to signal voicing distinctions, whereas in French they varied more in their reliance of VOT vs. onset f0, demonstrating the flexibility in the use of these correlates in the L2.

Keywords: individual production patterns, second language acquisition, French, English, voicing

1. INTRODUCTION

Recent studies indicate that variability in the acoustic realization of speech sounds is systematically constrained at the level of the individual speaker [5], [6], [17]. For example, previous research [5] has shown that within the same talker positive VOT of /p^h/ was highly correlated with that of /k^h/ (between-category covariation) in American English (AE). [6] also reports talker-based patterns of covariation in correlates of voicing (VOT, onset f0, vowel duration) across different realizations of /b/ and /p/. AE speakers have also been shown to maintain an inverse relationship between the amount of reliance on VOT vs. onset f0 in the individual realization of a voicing contrast [18]. Finally, [17] found a speaker-based negative correlation between the duration of prevoicing for /b/ and the duration of aspiration for /p^h/. Importantly, these studies investigated talker-based stability of phonetic realizations within a single language. If this stability is governed by an individual speaking style, or speech 'habit', as suggested by previous research it should be maintained across languages and within language in L2 learners [5].

The present study tests this hypothesis by examining the use of two correlates of voicing, Voice Onset Time (VOT) and onset f0 (fundamental frequency at the onset of voicing), in realization of voicing categories across English and French by intermediate level American learners of French. In particular, it is asked whether covariation between realizations of /b/ and /p/ exists on an individual level and whether it is maintained in both the first and second language. Evidence of such a relationship in the L2 would suggest that L2 speech categories are not acquired separately, but as a contrastive set.

In addition, it is asked whether members of phonological categories across languages are produced by each talker with similar phonetic settings. Evidence of such acoustic stability would suggest that L2 learners do consider, e.g., French /b/ and English /b/ to be phonologically equivalent and phonetically similar (for other work considering acquisition of L2 sounds as sets see [8]).

Finally, this study asks whether the individual pattern of relative reliance on multiple correlates of voicing is maintained across languages. In particular, this study explores the extent to which individual talkers employ VOT vs. onset f0 to construct voicing contrasts in each language and across languages.

Table 1: Voicing in French and English.

	[+voice]	[-voice]
French	Prevoiced (<0 ms)	Short lag (<30 ms)
English	Short lag (<30 ms) (some prevoiced)	Long lag (>30 ms)

The research questions above are based on the assumption that both VOT and onset f0 are important correlates of voicing in English and French [1], [10], [13]. The use of VOT to signal voicing in English and French is summarized in Table 1 [4], [12].

2. METHODS

2.1. Participants

Twenty-three native speakers of Midwestern American English (16 female and 7 male, mean age 19.96) participated in the study. At the time of the experiment, all participants were enrolled in at least one French course at the third semester level or above. The average length of French study across all participants was 7.11 years (ranging from 2-11 years) and average self-rated French proficiency was 4.85/7.

2.2. Stimuli

Stimuli consisted of four monomorphemic minimal pairs in each language, contrasting the voicing of the word-initial bilabial stop (e.g. bet/pet, bêche/pêche). Eight distractor pairs per language were also included (e.g. feed/deed, chaud/faux).

Vowels in both sets of stimuli were limited to front unrounded: /æ/, /ɛ/, /i/, and /ɪ/ in English and /i/, /ɛ/ and /a/ in French. Experimental items on both lists were frequent and familiar to both native and non-native speakers, as determined by [19], [15], and informal ratings by a native and a non-native speaker of French.

2.3. Procedure

Prior to recording, participants read a priming text in the language of the upcoming recording session. In order to control for speech rate, stimuli were presented one by one in randomized order on a screen using ePrime [16]. Participants were instructed to pronounce each word in their normal speaking voice. After completing three repetitions of the word list in each language, participants repeated the task in the other language starting with the priming text (the order of languages was counterbalanced). All recordings were conducted in a sound-attenuated booth using an ART Tube MP Project Series preamplifier and an Audio-Technica (AE4100) unidirectional cardioid dynamic microphone. Following the reading task, participants completed a language background questionnaire.

2.4. Measurements

Annotation and measurements were performed in Praat version 6.0.36 [3]. VOT was measured from the release of the stop to the onset of voicing. Onset f0 was measured at the first location following the stop release where the Praat autocorrelation algorithm could detect periodicity. To allow for cross-gender comparisons, onset f0 was semitone-normalized and values were examined for algorithm errors and manually corrected if necessary [9], [18]. VOT duration was not normalized since rate of speech was relatively consistent across and within participants. Burst duration of prevoiced stops was not measured given the relative scarcity of such realizations.

2.5. Analysis

The data was analysed primarily using a correlation. Discriminate analysis was used to estimate individual reliance on acoustic correlates of voicing in production. Results of the discriminant analysis were further examined with an ANOVA and correlation analyses.

3. RESULTS

3.1. Individual VOT of /b/ vs. /p/

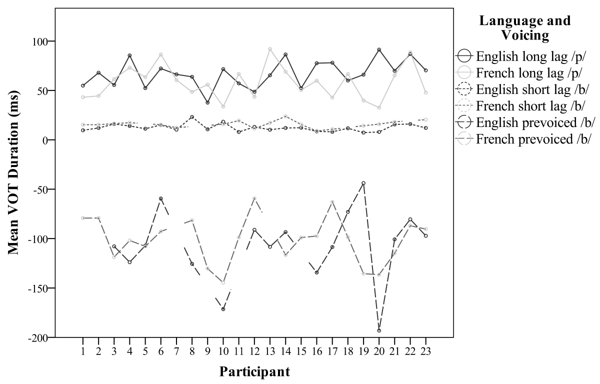
Table 2 reports the number (T#) and percentage (%) of tokens in each VOT category (30 ms was considered the boundary between short lag and long lag VOT). Fig. 1 plots individual mean VOT values for voicing categories in French and English (to avoid averaging across positive and negative values, short lag and prevoiced /b/ is shown separately) and demonstrates individual variability in these values.

Table 2: Number and percentage of VOT tokens.

	Prevoiced		Short lag		Long lag	
	%	T#	%	T#	%	T#
ENG	14.9%	82	36.1%	199	49.1	271
FRE	25.7%	142	31.2%	172	43.1%	238

Mean VOT values for both voicing categories in each individual speaker were tested for correlations across voicing categories. The results showed that the VOTs of /b/ and /p/ were not correlated on an individual level in English, whether the correlation was conducted between all voiced and all voiceless stops (r[23] = -.182, p=.407), between prevoiced /b/ and voiceless /p/ (r[23]= -.329, p = .197), or between short lag /b/ and voiceless /p/ (r[23] = .015, p = .945).

Figure 1: Individual mean VOT values for voicing in French and English.



Similarly, no significant correlations between VOTs of /p/ vs. /b/ were obtained in French: r[23] = .138, p = .55 (prevoiced /b/ vs. /p/).

Overall, the results indicate that acoustic realization of one member of the contrast (e.g. /b/) was not related to the acoustic realization of another member (/p/), in either English or French.

3.2. Correlations between individual production in English and French

Mean individual VOT values for each voicing category in English were correlated with mean values in French. Since voiced stops could be realized as two distinct VOT types in both English and French – prevoiced or short lag – they were considered separately. The results showed that VOTs of French and English /p/ were not correlated (r[23] = .137, p = .533). Voiced stops produced with positive VOT were also not correlated across languages (r[18] = .08, p = .753)

Figure 2: Correlation between French and English prevoicing duration.

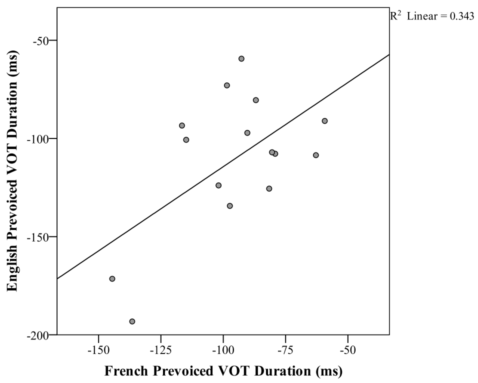
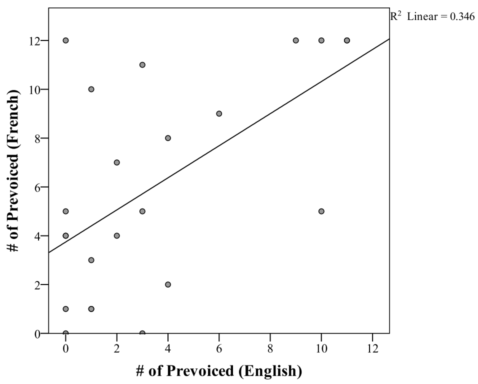


Figure 3: Correlation between number of instances of prevoicing in English and French.



However, duration of prevoicing in English /b/ (for those English participants who prevoiced) was significantly positively correlated with duration of prevoicing for French /b/: r[15] = .586, p = .022 (one outlier with the VOT value more than 1.5 standard deviations away from the mean was removed for this analysis). This result suggest that participants were more successful at producing longer prevoicing in French if their English prevoicing values were relatively long (Fig. 2).

Learning to produce French /b/ does not involve shortening or lengthening the VOT of English /b/ (predominantly, produced with positive short lag VOT) but rather switching one's production to another VOT type – negative VOT. Therefore, the incidence of correct VOT type production in French is as important as VOT duration. To investigate the possible link between French and English incidences of prevoicing, the number of prevoiced tokens produced by each participant in each language was calculated. The analysis showed that there was a significant positive correlation between number of prevoiced /b/s in English and number of prevoiced /b/s in French (r[23] = .588, p = .003, Fig. 3). Participants who were prevoicing more frequently in English were more likely to prevoice in French.

3.3. Using VOT and onset f0 in realizations of voicing contrast across English and French

To establish the relative weighting of VOT and onset f0 in the individual production of voicing contrast in both languages, a discriminant analysis was deployed [7]. It generated a set of individual standardized canonical coefficients for each correlate of voicing. These coefficients are indicative of how much each speaker relied on each of the correlates in realizing the distinction between voiced and voiceless consonants. Individual coefficients were tested for the effects of Language (English vs. French) and Correlate (VOT vs. onset f0). Results showed that across languages VOT was assigned a significantly greater weight than onset f0 in differentiating voicing categories (effect of Correlate in a repeated measures ANOVA: F(1, 22) = 78.596, p < .001). There was no significant effect of Language or Language by Correlate interaction. Thus, in both languages, participants as a group relied more on VOT than onset f0 in realizing the voicing contrast.

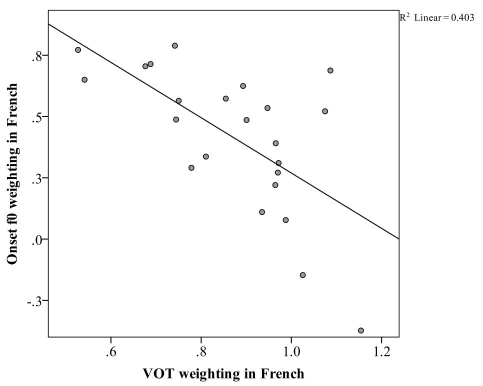
The coefficients were then checked for within and across-language correlations. There was no correlation between VOT weights and onset f0 weights in English (r[23] = -.251, p = .248). Overall, in English a relatively wide range of onset f0 weights was associated with a considerably more restricted distribution of VOT weights.

In French, a significant negative correlation was present between VOT and onset f0 weights: r[23] = -.635, p = .001. Participants who relied on VOT less assigned a greater weight to onset f0 in realizing the contrast between voiced and voiceless stops (Fig. 4).

In contrast to English, a comparable amount of variability was observed for individual VOT and onset f0 weights in French.

Participants weighting of VOT in English was uncorrelated with their weighting of VOT in French (r[23] = -.030, p = .893. The same lack of crosslinguistic correlation was observed for onset f0 weight (r[23] = 0.0003, p = .999). Thus, participants who tended to rely strongly on a given correlate in English did not necessarily demonstrate a comparable degree of reliance in French.

Figure 4: Correlation between VOT and onset f0 weights in realization of French voicing contrast.



4. DISCUSSION

The results of this study did not show a correlation between phonetic realizations of /b/ and /p/ in either French or English speech. In a higher-powered study, [5] also report a considerably weaker relationship within homorganic pairs (/b/ and /p/) than across heterorganic stops (/p/, /t/, /k/). The lack of covariation in French may suggest that members of the voicing opposition are acquired separately in the L2, thus not exhibiting mutual effects on each other's realization. In other words, learning to realize one member of the contrast in L2 in a more native-like fashion does not presuppose that another member will be proportionately successful. However, this relationship was also absent in the L1, which may indicate that considerations of contrast maintenance are not of great importance in guiding phonetic realization. Finally, it is also possible that this relationship did not reach significance in the current study due to insufficient statistical power (cf. [5]).

Across languages, only duration of prevoicing showed a moderate correlation. Realization of voiceless stops (largely produced as aspirated in both English and French) and of non-prevoiced voiced stops were not correlated across languages. The strongest correlation was obtained between the instances of prevoicing for /b/ across the two languages. This suggests that among voicing categories participants perceived the greatest affinity between English and French prevoiced /b/. This is perhaps not surprising given that it is the only category with the same phonological status and the possibility of identical phonetic realization across the two languages. Interestingly, although voiceless stops were also (erroneously) realized in a largely identical manner across languages, their VOTs were not correlated. This could be interpreted as an indication that their cognitive representations were distinct, although their acoustic realizations converged. Ultimately, evidence for the crosslinguistic link in realization of voicing categories is fairly tentative. The results suggest that the 'habit' of producing English /b/ with prevoicing helps L2 learners to target French prevoiced /b/ with a greater success. Alternatively, those participants who were successful at learning French prevoicing may have transferred this 'habit' to English.

With respect to the relative reliance on VOT vs. onset f0 in signalling voicing distinctions, participants uniformly relied on VOT to a greater extent. This was particularly apparent in their English productions, where VOT was consistently assigned a great discriminant weight with little variability across individuals. There was also no connection between VOT and onset f0 reliance in English. In French, the picture was considerably different. Participants varied to a greater extent in their reliance on VOT when constructing an acoustic distinction between /b/ and /p/ in French. This could result from learners' shortening of voiceless VOT (aiming for French short lag targets) which was not sufficiently compensated by shifting voiced productions into the negative VOT region (see Fig. 1). As a result, VOT becomes less distinctive as a correlate of voicing. Interestingly, there was a negative correlation between reliance on VOT and onset f0 in French, suggesting that a decreased distinctiveness of VOT was compensated by a relatively exaggerated reliance on onset f0, on an individual basis. The fact the participants maintained this compensatory relationship in their L2 but not L1 suggests flexibility in deployment of multiple acoustic cues signalling the distinction. Given this finding, it is plausible that speakers apply similar strategies when distinctiveness of the primary correlate is reduced due to other reasons, such as noisy environment, reduction in rapid speech or in highly frequent words [2], [11], [12]. Such an ability to compensate with secondary correlates in compromised settings would mirror an equivalent behaviour established in speech perception: secondary cues to the contrast exert a greater influence on speech categorization when primary cues are obscured or unavailable [20].

7. REFERENCES

[1] Abramson A. S., Lisker, L. 1965. Voice onset time in stop consonants: Acoustic analysis and synthesis. In *Proc. 5th international congress of acoustics* (Vol. 51). A51, Liege.

[2] Bell, A., Breiner, J. M., Gregory, M., Girand, C., Jurafsky, D. 2009. Predictability effects on duration of content and function words in conversational English. *Journal of Memory and Language.* 60(1), 92-111.

[3] Boersma, P., Weenink, D. 2009. Praat: Doing phonetics by computer (Version 6.0.36) [Computer program]. Amsterdam, The Netherlands: University of Amsterdam.

[4] Caramazza, A., Yeni-Komshian, G.H. 2005. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. The *J. Acoust. Soc. Am*. 54(2), 421-428.

[5] Chodroff, E., Wilson, C. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English. *J. Phon.*, 61, 30-47

[6] Clayards, M. 2017. Individual talker and token covariation in the production of multiple cues to stop voicing. *Phonetica.* 75(1), 1-23.

[7] Cohen et al. 2003. Applied multiple regression/correlation analysis for the behavioral sciences. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(4) 691-691.

[8] De Jong, K. J. Silbert, N. H., & Park, H. 2009. Generalization across segments in second language consonant identification. *Language learning*, *59*(1), 1-31.

[9] Dmitrieva, O., Llanos, F., Shultz, A. A., Francis, A. L. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *J. Phon.*, 49, 77-95.

[10] House, A. S., Fairbanks, G. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. Acoust. Soc. Am*. 25, 105-113.

[11] Kessinger, R. H., Blumstein, S. E. 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. *J. Phon.,* 25(2), 143-168.

[12] Lisker, L., Abramson, A.S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.

[13] Ohde, R. 1984. Fundamental frequency as an acoustic correlate of stop consonant voicing. *J. Acoust. Soc. Am*. 75, 224-240.

[14] Miller, J. L. 1981. Effects of speaking rate on segmental distinctions. *Perspectives on the study of speech.* 39-71.

[15] New, B., Pallier, C., Ferrand, L., Matos, R. 2001. Une base de données lexicales du français contemporain sur internet: LEXIQUE, *L'Année Psychologique*, 101, 447-462. http://www.lexique.org

[16] Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-Prime User's Guide (Psychology Software Tools Inc., Pittsburg, PA).

[17] Scobbie, J. M. 2006. Flexibility in the face of incompatible English VOT systems. In L. Goldstein, D. H. Whalen, C. T. Best (Eds.), *Laboratory phonology 8: Varieties of phonological competence.* New Haven, Conn.

[18] Shultz, A. A., Francis, A. L., Llanos, F. 2012. Differential cue weighting in perception and production of consonant voicing. *J. Acoust. Soc. Am*, 132(2). 95-101.

[19] Washington University in St. Louis, Speech & Hearing Lab Neighborhood Database. Available from http://128.252.27.56/Neighborhood/SearchHome.asp

[20] Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. 1990. Gradient effects of fundamental frequency on stop consonant voicing judgements. *Phonetica*, 47(1-2), 36-49.