TOWARDS AN ACCOUNT OF LANGUAGE-SPECIFIC PATTERNS OF THE TIMING OF VOICING

C. Wills and G. Docherty

Department of Speech, University of Newcastle-upon-Tyne, UK.

ABSTRACT

Results are presented from a study of the timing of voicing in English obstruents produced by native speakers of French and Spanish. It is suggested that in attempting to account for the timing of voicing (in native as well as non-native performance) an incomplete picture may be obtained if the variability in speaker performance is omitted from the account, and if excessive attention is focused on VOT as opposed to overall laryngeal-supralaryngeal coordination.

1. INTRODUCTION

Instrumental studies of the phonetic performance of non-native speakers of a language (particularly English) have been used (a) as evidence in support of a model of acquisition of L2 (e.g.[2]), and (b) to shed light on the status of fine-grained phonetic variability, specifically regarding the extent to which it is a language-specific and learned aspect of phonetic performance (e.g. [3,6]). Studies of consonant production in L2 have focused almost exclusively on VOT, and the basis of comparison between different groups of subjects has typically been the mean VOT for particular categories of stops.

In this communication it is suggested that by commonly adopting the approach just described, previous studies may be overlooking some of the fundamental characteristics of the L2 (and L1) speaker performance. It has recently been proposed that the phonetic representation of an utterance may consist not of a string of precise target specifications, but may instead be characterised by built-in variability and underspecification [1,4]. The implication of this is that an account of performance focusing exclusively on mean scores may only be painting part of the picture. Furthermore, recent work on both the detailed characteristics of laryngeal timing in stops [5] and on the phonetic and phonological representation of the voicing contrast [1] suggests that greater observational and explanatory emphasis should be placed on the overall timing and coordination of laryngeal and supralaryngeal gestures, and that variability of VOT (for example) may arise from variability of other timing and control parameters as opposed to being directly manipulated itself.

It seems timely therefore to investigate whether these revised notions of target and control with regard to the timing of voicing lead to rather different inferences being made about speech production from data obtained from L2 performance. This is the aim of a project being undertaken at the University of Newcastle-upon-Tyne, and the goal of this paper is to present a snapshot of some early results.

2. PROCEDURE

The aim of the experiment described below is to study the production of /p/, /b/, /s/ and /z/ in English by native speakers of French and Spanish. This paper deals only with the results pertaining to /b/ and /z/. Five native speakers of French and three of Spanish were recruited. The French speakers had all lived in the North-East of England for over 8 years. Due to difficulties in locating subjects, the group of Spanish speakers was rather heterogeneous (a factor to be borne in mind in interpreting the results). SP1 had lived in the UK for 2 years, SP2 for 20 years, SP3 for 15 years. A group of 5 native English speakers was used for control purposes. Henceforth the subjects are referred to as ENG1-5, FR1-5 and SP1-3.

All the speakers were recorded producing (a) a list of 16 isolated English words (5 repetitions) containing 4 cases each of initial /p,b,s,z/; (b) the same words embedded in a carrier sentence (5 repetitions). The FR and SP speakers were also recorded producing 5 repetitions of a matched set of isolated (16) French and (12) Spanish words respectively (the Spanish list was shorter due to absence of initial /z/ in Spanish) and the same words embedded in a French or Spanish carrier sentence (only 3 repetitions of the carrier sentences were obtained from SP1-3). The conditions are referred to henceforth as (1) Eng/Eng (i.e. English subjects/English words or sentences) (2) Fr/Fr (3) Fr/Eng (4) Sp/Sp (5) Sp/Eng. High quality DAT recordings were made in studio conditions. Subjects were asked to read the material from printed lists at a comfortable rate.

Wide-band spectrograms were made of the data using a LSI Speech Workstation, and were displayed on the screen of a PC terminal aligned with the corresponding speech waveform. The following measurements were taken for each token: VOT (stops only) taken as the interval between the release burst of a stop and the onset of the first vertical striation for a following vowel; stop duration defined as the interval between the release burst of the stop and the point at which the second and higher formants disappeared from the spectrogram during the transition from the preceding vowel (this measurement could only be performed in the carrier sentence conditions given the need for a preceding vowel context); fricative duration defined as the interval between the onset and offset of the noise component visible in the spectrogram corresponding to the fricative; medial voicing, defined as the presence of vertical striations during the intervals previously identified as a stop or fricative.

3. RESULTS

Space prevents a detailed exposition of the results. Table 1 shows the mean VOT, consonant duration and medial voicing for /b/ and /z/ produced under the various conditions by the FR and SP speakers only. The principal findings are as follows.

- In /b/ in isolated words both negative and positive VOTs are found in Fr/Fr, Fr/Eng and Eng/Eng. Eng/Eng stops have few negative VOTs, Fr/Eng rather more, and Fr/Fr the most. The results for Sp/Eng speakers differ according to the subject. SP1 produced only prevoiced (i.e. negative VOT) stops in both languages. SP2 produces both short lag VOT and prevoiced stops in both languages, but with a difference in weighting such that the VOTs are predominantly short lag in the English words and negative in the Spanish words. SP3 uses both patterns in English without any apparent weighting, but produced almost exclusively prevoiced stops in Spanish.

- In /b/ in carrier sentences, both the English and French subjects' performance is characterised by a good deal of variability. On the whole, the Fr/Fr stops are more 'voiced' (i.e. more commonly entirely voiced, and with generally proportionally longer intervals of medial voicing) than stops produced in either the Fr/Eng or Eng/Eng conditions. The stops in the latter two conditions are similar with the exception that the Fr/Eng stops are considerably longer on average. There are large differences in the realisation of /b/ by SP subjects across the two languages. SP1 and SP3 in the Sp/Sp condition produce /b/ predominantly as fully voiced labial approximants. In the Sp/Eng conditions, both subjects consistently produce stop closures, but with variable timing of voicing, producing both fully voiced and partially devoiced tokens of /b/. SP2 produces both fully voiced and partially devoiced stops across all three conditions.

- in /z/ in isolated words the principal feature is the variability in the data. French and English subjects produce predominantly fully voiced or partially

Table 1: Mean VOT, consonant duration (CD), medial voicing (MV), and no. of cases of (A) complete devoicing, (B) partial devoicing and (C) complete voicing observed in /b/ and /z/ by FR and SP speakers (figures in parentheses are standard deviations/number of cases). All means and s.d.s are given in ms. Shortfalls in *ns* reflect cases where either speaker error or measurement uncertainty forced exclusion of a token.

Mean VOT in /b/ in isolated words -- separate means given for +ve and -ve VOTs)

Fr/Fr			Fr/Eng		
Fr1	19(8/5)	-71(20/13)	Fr1	21(6-17)	-74(-/1)
Fr2	13(5/5)	-72(14/5)	Fr2	11(2/13)	-96(31/6)
Fr3	24(15/16)	-44(24/4)	Fr3	16(4/18)	-74(-/1)
Fr4	17(-/1)	-80(22/15)	Fr4	11(2/7)	-72(42/8)
Fr5	12(3/8)	-62(22/9)	Fr5	-	-124(12/6)
Sp/Sp			Sp/Eng		
Sp1	-	-119(20/20)	Sp1	-	-119(23/20)
Sp2	14(3/4)	-52(10/15)	Sp2	14(3/17)	-57(11/3)
Sp3	10(-/1)	-48(21/17)	Sp3	26(22/9)	-65(32/11)

Mean consonant duration, medial voicing and summary of timing patterns for /b/ in carrier sentences

Fr/Fr	CD	MV	A	B	C	Fr/Eng(CD)	MV	A	B	C
Fr1	74(14/20)	74(14/20)	-	-	20	151(84/20)	137(40/19)	1	1	18
Fr2	91(30/20)	90(18/20)	-	3	17	166(70/20)	140(43/17)	3	6	11
Fr3	67(7/20)	61(14/20)	-	9	11	99(15/20)	83(19/20)	-	7	13
Fr4	103(22/20)	100(24/20)	-	3	17	186(69/20)	98(28/20)	-	16	4
Fr5	93(21/19)	71(29/19)	-	11	8	220(79/16)	110(57/16)	0	10	6
Sp/Sp						Sp/Eng				
Sp1	-	-	-	-	-	113(26/11)	92(61/7)	4	3	4
Sp2	61(14/11)	54(14/11)	-	4	7	100(18/9)	60(21/9)	-	7	2
Sp3	66(19/6)	66(19/6)	-	-	6	102(20/7)	85(26/7)	-	3	4

Mean consonant duration, medial voicing and summary of timing patterns for /z/ in isolated words

Fr/Fr	CD	MV	A	B	C	Fr/Eng(CD)	MV	A	B	C
Fr1	140(51/20)	146(55/14)	6	2	12	168(43/20)	163(38/10)	10	1	9
Fr2	124(26/20)	98(46/16)	4	8	8	140(35/20)	109(64/18)	2	9	9
Fr3	122(18/20)	66(47/19)	1	11	8	124(14/20)	49(40/18)	2	14	4
Fr4	134(29/20)	43(32/20)	-	19	1	133(37/20)	38(43/16)	4	15	1
Fr5	150(37/20)	158(34/17)	3	-	17	142(48/19)	134(60/19)	-	2	17
Sp/Sp						Sp/Eng				
Sp1	-	-	-	-	-	124(50/20)	106(66/13)	7	3	10
Sp2	-	-	-	-	-	60(13/20)	56(17/16)	4	1	15
Sp3	-	-	-	-	-	128(25/20)	87(62/8)	12	4	4

Mean consonant duration, medial voicing and summary of timing patterns for /z/ in carrier sentences

Fr/Fr	CD	MV	A	B	C	Fr/Eng(CD)	MV	A	B	C
Fr1	140(42/20)	140(42/20)	-	-	20	165(30/20)	159(40/16)	4	1	15
Fr2	141(35/20)	131(43/20)	-	3	17	160(37/20)	146(49/19)	1	3	16
Fr3	107(12/20)	101(23/20)	-	6	14	121(15/20)	96(40/20)	-	6	14
Fr4	154(40/20)	128(52/20)	-	9	11	135(20/20)	116(40/17)	3	13	4
Fr5	124(27/20)	96(41/20)	-	9	11	188(37/20)	173(55/20)	-	14	6
Sp/Sp						Sp/Eng				
Sp1	-	-	-	-	-	124(18/12)	56(-/1)	10	1	1
Sp2	-	-	-	-	-	84(14/11)	76(21/10)	1	2	8
Sp3	-	-	-	-	-	148(9/12)	48(22/9)	3	9	-

devoiced /z/ in all three conditions. No trends emerge regarding changes produced by French speakers in their performance of Fr/Eng. Two features emerge from the Spanish data (in which, of course, speakers face a novel situation given the absence of word-initial /z/ in Spanish). SP1 consistently initiates phonation before the start of the fricative noise characterising the /z/ (mean voicing lead = 121ms -- not shown in Table 1), and on some occasions proceeds to produce a fricative without any phonation, whilst on others voicing continues all the way through the fricative into the following vowel. In all the SP subjects there is a tendency for there to be a larger number of cases of completely devoiced /z/ in the Sp/Eng condition than in the Eng/Eng condition.

- in /z/ in the carrier sentence condition, variability in realisation is the principal feature, with cases of full voicing and partial devoicing being found across all the subjects, and with full devoicing being found somewhat less frequently. In the Fr/Eng condition, there is a tendency for /z/ to have shorter intervals of medial voicing than are found in Fr/Fr tokens. Completely voiceless tokens of /z/ are found more commonly in Sp/Eng than in Eng/Eng.

4. DISCUSSION

The absence of data from monolingual French and Spanish speakers precludes at this stage a statement regarding the degree of interaction of L1/L2 in the data (e.g. along the lines described in [2]), but the results do confirm the findings of (amongst others) [3] and [6] that the fine detail of phonetic realisation may be altered in the production of consonants in L2. The results conform to previous studies showing that VOT is one parameter which can be observed to alter in L2 performance. The data pertaining to /z/ shows that speakers are also able to manipulate laryngeal-supralaryngeal timing in the production of other sounds. In the light of this, it would seem fruitful to work towards a broader account of this aspect of non-native speaker performance than has been offered so far, recognising that VOT is a reflection of a more general process of gesture coordination, and thereby approximate an account which covers timing of voicing in general as opposed to only in stops.

The data also suggests that an account of the speakers' performance which presented no more than the mean VOT, consonant duration, and medial voicing would only paint part of the picture, and in particular would obscure the abundant inter- and intra-subject variability observed in the data, and consequently one of the major features of that data. For example, the mean medial voicing for FR1's /z/ in Fr/Eng isolated words would not be a good reflection of the fact that half of the tokens produced by FR1 are completely devoiced, and almost all the remainder are fully voiced. The observations made in this study could only be *fully* characterised by consideration of means *and* some measure of variance. Observations expressed in this way will allow full evaluation of the subjects' performance in the light of the work mentioned in 1. regarding inherent variability in phonetic targeting. This work is now in progress.

5. REFERENCES

[1] DOCHERTY G. J. (1991) *The Timing of Voicing in British English Obstruents.* Dordrecht:Foris

[2] FLEGE, J. (1991) "Age of learning affects the authenticity of voice-onset time in stop consonants produced in a second language" *J.A.S.A., 89*:395-411

[3] FLEGE, J. & PORT, R. (1981) "Cross-language phonetic interference: Arabic-to-English" *Lg. & Sp., 24*:125-146

[4] KEATING, P. (1988) "The window model of coarticulation" *UCLA WPP, 66*:104-123

[5] LOFQVIST, A., & YOSHIOKA, H. (1984) "Intrasegmental timing: laryngeal-oral coordination in voiceless consonant production" *Speech. Comm., 3*:279-289

[6] PORT, R. & MITLEB, F. (1980) "Phonetic and phonological manifestations of the voicing contrast in Arabic-accented English" *Research in Phonetics, 1*:137-165 (Dept. of Linguistics, Indiana University).