Factors affecting the perception of plosives in second language English by first language Cypriot-Greek listeners

Kkese, Elena[1], Petinou, Kakia[2]
ElenaKkese@hotmail.com, kakia.petinou@cut.ac.cy

[1]Cyprus University of Technology, UCLan Cyprus, [2]Cyprus University of Technology

Abstract. This paper investigates the difficulties adult second language users of English encounter with plosive consonants. It presents the results of a word identification task examining the acquisition of plosive voicing contrasts by college students with Cypriot-Greek background. The task by using minimal pair words focused on investigating possible factors affecting plosive identification. Both descriptive and inferential analyses were used for identifying how important each factor is when it comes to plosive consonants. The results provide an indication of the rank order for the examined factors. Specifically, syllable position is identified as having the greatest influence on plosive identification, followed by voicing, word position, and place of articulation for both kinds of analyses. By accepting the hypothesis that less successful differentiation of plosive consonants in the second language on the part of Cypriot-Greek users was partially due to the investigated factors manifested implies that the specific second language sounds do not exist or are non-contrastive in the first language. Nonetheless, because the weighting of auditory cues in the categorisation of plosives is language-specific, participants were modifying their identification of voiced plosives to fit the mother tongue. Speech perception can, therefore, account for the data of the present study. Specifically, Voice Onset Time provides important information to the voiceless-voiced distinction as well as word-initial, syllable onset plosive consonants.Taken together, the results of the present study indicate that when dealing with contrastive categories in the second language, the acoustic cue of Voice Onset Time is of crucial importance. For Cypriot-Greek users, acquiring voiced plosives means acquiring new Voice Onset Time patterns.

Keywords: second language, speech perception, plosive consonants, voice onset time

Introduction

Plosive consonants in English and CG

A comparison of the plosive system of English and Cypriot-Greek (CG) indicates that both languages have plosive consonants at three places of articulation: bilabial, alveolar and velar. English and CG plosive consonants, however, differ in the number of plosives and their acoustic realisations. These differences may actually be the reason for the difficulties of second language (L2) users when attempting to acquire the L2 plosive system.

Voice Onset Time (VOT) is normally used to characterise plosive consonants across languages. VOT, the period between the plosive closure release and the beginning of voicing, is the primary acoustic cue for the voicing distinction (Lisker & Abramson, 1964). Three patterns for VOT production are evident (Lisker & Abramson, 1964) involving the long voicing lead in which phonation begins before the oral release (voiced plosives), the short voicing lag in which phonation begins just after the oral release (voiceless unaspirated plosives), and the long voicing lag in which phonation begins after the oral release (voiceless aspirated plosives). In English, plosives can be produced with a long voicing lead and a short voicing lag (voiced *vs.* voiceless) (Okalidou et al., 2010). Therefore, English is a two-category language consisting of voiced and voiceless plosive consonants. Specifically, VOT is considerably longer for voiceless plosives than voiced ones yielding approximate mean values of 58ms for [p] versus 1ms for [b], 70ms for [t] versus 5ms for [d], and 80ms for [k] versus 21ms for [g] (Lisker & Abramson, 1964).

In CG, though, the situation is more complex regarding plosive consonants. The two views proposed for the plosive system agree on the presence of short versus long lag times but there is disagreement whether voiced plosives are contrastive segments. With reference to VOT, this is considerably longer for voiceless aspirated plosives compared to voiceless unaspirated ones with mean values of 55ms for [pʰ] versus 5ms for [p], 60ms for [tʰ] versus 17ms for [t], and 65ms for [kʰ] versus 22ms for [k]

(Tserdanelis & Arvaniti, 2001; Arvaniti, 1999). This suggests that CG voiceless plosives differ from the English ones while voiceless aspirated plosives are the ones that are closer to the English voiceless ones. With reference to voiced plosives, descriptions vary considerably. According to Arvaniti (2010), CG consists of unaspirated and aspirated voiceless plosives while it has no voiced plosives. This explains why English words such as 'league' [liːg] may be pronounced as [lik]. Further descriptions maintain that voiced plosives do exist in CG (Okalidou et al., 2010; Botinis et al., 2004). Based on these accounts, the plosive consonants can be divided into three voicing categories, namely, voiced unaspirated plosives, voiceless unaspirated plosives, and voiceless aspirated plosives. Regarding voiced unaspirated plosives, they are always found in a post-nasal position as in [kuˈ^{m}bin] (button). The second and third categories are manifested in the minimal pair [kuˈpin] (oar) and [kuˈp^{h}ːin] (small bowl).

Method

Research Question

Through descriptive and inferential analyses, the study attempted to answer the question:

What are the factors that affect the identification of plosive consonants for CG users of L2 English?

Specifically:

a. What is the effect of consonant voicing in the identification of plosives?

b. How is the identification of plosives influenced by their respective place of articulation?

c. In which word position(s) (word-initial, -medial, -final) are plosives most easily identified?

d. In which position in a syllable are plosives more easily identified (onset/coda)?

The research approach used was quantitative aiming at identifying the factors affecting plosive identification. Differences were examined in the dependent variable (percentage of correctness) thought to be caused by the independent variables (four aforementioned factors).

Word identification task

For the developed task, a total of 120 target items were compiled that were arranged in 60 minimal pairs focusing on the voicing contrast of plosive consonants. Low-frequency words were preferred because they cannot be identified on the basis of fewer perceptual features. Nonetheless, the words had transparent spelling. Each pair of words was parallel in distribution and semantically contrastive differing in only one sound that could be found word-initially, -medially, or -finally (*i.e.* word-initial: palate – ballot, tessellated – desolated, crypt – gripped; word-medial: apace – abase, mettlesome – meddlesome, lacquered – laggard; word-final: grippe – grebe, alight – allied, burke – berg).

16 minimal pair targets were included for each category of consonants while distractors focusing on the voicing contrast were also intermixed and made up 12 of the minimal pair words. Specifically, 2 to 4 distractors were used for every 16 presentations. These included fricative consonants such as the labiodental [f] and [v], dental [θ] and [ð], and alveolar [s] and [z], while some involved the palato-alveolar affricate consonants [ʧ] and [ʤ]. The words were presented in two fully randomised blocks in order to exclude any systematic patterning while two versions of this task were created in which the selection of items was entirely complementary. This type of task was chosen to eliminate any semantic information from the input (context-free) as it would have happened if a conversation was presented instead.

Participants and procedure

113 CG users of L2 English with typical speech and hearing were recruited for the purpose of this study. In order to ensure a homogenous participant pool (introductory level students), the participants were first-year students in an English-speaking college. In that way, participants shared the same characteristics (*e.g.* first language, educational level, socio-economic status) since the aim was to eliminate inter-group differences. Participation in the task was on a completely voluntary basis and students were ensured about the confidentiality of their personal detail. The only cases in which participants were excluded from the sample involved students whose first language (L1) was not CG.

The research period involved three spring semesters in order to investigate whether different students of the same level and background face the same difficulties with the specific sounds. The task was pre-recorded using Audacity 1.3 Beta software for recording and editing sounds. The speaker (one woman, age 30) was a native speaker of RP (Received Pronunciation). She was told to read at her normal pace without any particular attention to clarity and to imagine that the intended listeners were highly familiar with her voice. The task was administered as a two-alternative forced-choice task via a circling response mode. On each trial, participants listened to a target word along with its foil and responded by circling the word heard. Responses were scored as correct or incorrect generating an overall percent correct score as well as percent correct scores for the feature classes of consonant voicing, place of articulation, word position, and syllable position. Participants had the opportunity to listen to the words twice while response order and stimulus were counterbalanced.

Results

Quantitative findings: factors affecting plosive identification

Since two different formats of the task were administered, independent sample t-tests were conducted to compare mean performance overall. The performance indicated that there was not a significant difference between the two versions of the task, which were administered to the two groups ($p > .05$).

Effect of voicing

A MANOVA (Multivariate Analysis of Variance) with three articulation levels (bilabial, alveolar, velar) and two voicing levels (voiceless, voiced) as within subject factors was conducted in order to compare effects of voicing and place of articulation on performance. The MANOVA indicated a significant main effect of voicing on performance, $F(1,112) = 16.50$, $p < .001$, $\eta^2 = .13$ with overall performance (% correct answers) being higher for voiceless plosive consonants ($M = 56.33$, $SE = 2.15$) than voiced ($M = 49.85$, $SE = 1.90$). There were no significant main effects for place of articulation but the interaction between voicing X place of articulation was significant $F(2,224) = 7.28$, $p < .001$, $\eta^2 = .06$, such that participants performed considerably better in perceiving voiceless plosives only when place of articulation was bilabial (for [p] % correct $M = 58.41$, $SD = 28.66$; for [b] $M = 45.65$, $SD = 23.36$) or velar (for [k] % correct $M = 57.23$, $SD = 25.85$; for [g] $M = 50.57$, $SD = 25.27$) but not alveolar (for [t] % correct $M = 53.36$, $SD = 26.89$; for [d] $M = 53.33$, $SD = 26.03$). Specifically, there was an advantage of bilabial (voiceless $M = 58.41$, $SE = 2.70$; voiced $M = 45.65$, $SE = 2.20$) and velar plosives (voiceless $M = 57.23$, $SE = 2.43$; voiced $M = 50.57$, $SE = 2.38$) compared to alveolar (voiceless $M = 53.36$, $SE = 2.53$; voiced $M = 53.33$, $SE = 2.45$). The interaction is indicated in Figure 1.

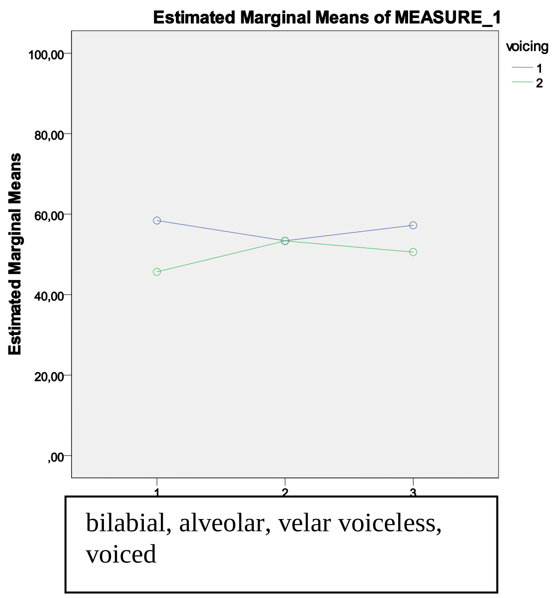


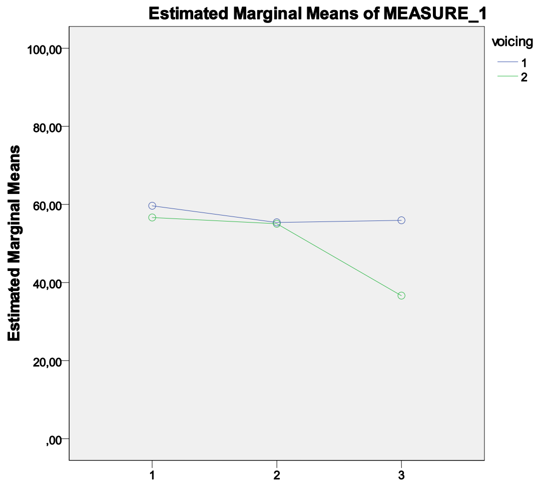
Figure 1. Interaction between voicing X place of articulation (better in bilabial and velar voiceless plosives)

Effect of word position

The second MANOVA (voicing X place of articulation X word position) focused on word position as a function of place of articulation and voicing. Significant main effects for voicing (better performance for voiceless), $F(1,112) = 21.60$, $p < .001$, $\eta^2 = .16$ were maintained and an additional main effect for word position was identified, $F(1.8,202.46) = 13.54$, $p < .001$, $\eta^2 = .11$ (Greenhouse-Geisser corrected). For word position, the effect was significant for all three levels with best performance for initial word position ($M = 58.14$, $SE = 2.89$) compared to both medial ($M = 55.22$, $SE = 2.35$) and final ($M = 46.28$, $SE = 1.46$) and with best performance for medial word position compared to final. Especially, the advantage for voiceless ($M = 56.99$, $SE = 2.10$) over voiced plosives ($M = 49.44$, $SE = 1.95$) was greatest for final word position, less pronounced for initial word position, and absent for medial word position. The interaction between word position X voicing was significant $F(2,224) = 18.46$, $p < .001$, $\eta^2 = .14$. Specifically, participants performed significantly better in perceiving voiceless consonants in all three categories (word initial $M = 59.64$, $SE = 3.01$; word-medial $M = 55.39$, $SE = 2.79$; word-final $M = 55.92$, $SE = 2.09$) compared to their voiced counterparts (word initial $M = 56.64$, $SE = 3.14$; word-medial $M = 55.05$, $SE = 2.55$; word-final $M = 36.63$, $SE = 2.01$). Contrasts revealed that the percentage of correct responses was significantly higher for medial $F(1,112) = 14.43$ and initial word position $F(1,112) = 19.31$ compared to the final word position, *ps* $< .001$. However, effect of voicing (advantage for voiceless over voiced) was most pronounced for the word-final category (Figure 2).

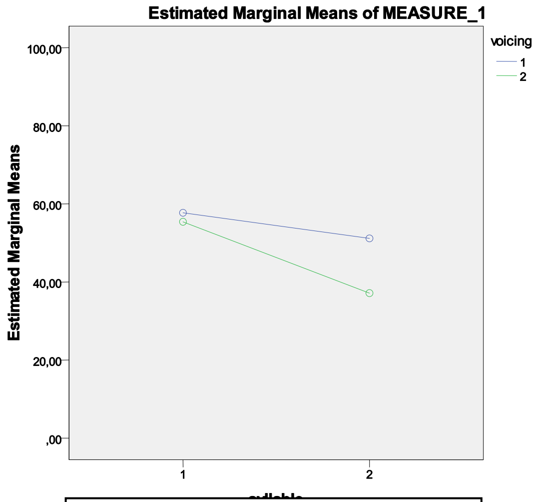
Effect of syllable position

The third MANOVA (voicing X place of articulation X syllable) indicated significant main effects for all variables: voicing $F(1,112) = 20.76$, $p < .001$, $\eta^2 = .16$, place of articulation $F(2,224) = 3.66$, $p < .05$, $\eta^2 = .03$, and syllable $F(1,112) = 30.12$, $p < .001$, $\eta^2 = .21$. Concerning syllable position, performance was significantly better for onset ($M = 56.57$, $SE = 2.46$) compared to coda ($M = 44.15$, $SE = 1.29$), especially for bilabial consonants (for [p] % correct $M = 60.23$, $SD = 33.54$; for [b] $M = 53.83$, $SD = 32.87$) compared to alveolar (for [t] % correct $M = 53.51$, $SD = 32.11$; for [d] $M = 57.20$, $SD = 30.90$) and velar consonants (for [k] % correct $M = 59.42$, $SD = 31.59$; for [g] $M = 55.21$, $SD = 31.82$). A significant interaction was identified between syllable X voicing $F(1,112) = 17.33$ $p < .001$, $\eta^2 = .13$, which indicated that participants did better in voiceless consonants especially at onset position (voiceless M = 57.72, SE = 2.65; voiced M = 55.42, SE = 2.56) compared to coda (voiceless M = 51.16, SE = 1.72; voiced M = 37.14, SE = 2.02). However, effect of voicing (advantage for voiceless over voiced) was most pronounced for coda (Figure 3).



word-initial, word-medial, word-final voiceless, voiced

Figure 2. Interaction between voicing X word position (better in voiceless plosives in word-initial position)



onset, coda voiceless, voiced

Figure 3. Interaction between voicing X syllable position (better in voiceless onset plosives)

Rank Order of the factors affecting plosive consonant identification

Descriptive statistics combining information about all investigated factors regarding the word identification task involving voicing, place of articulation, word position, and syllable position was used to determine frequency of factors. From a descriptive look, by ranking the means for all combinations in descending order (see table 1), it seems that phonemes in the first (best performance) rows of this table tend to be the ones in onset syllable position. Specifically, the actual difference between scores tends to be greater as a function of syllable position rather than a function of the other factors. Voicing seems to have the second greatest influence on plosive consonant identification since voiceless have higher scores while the ones with the lowest tend to be voiced. Word position follows and is identified as the third most important factor while the last factor identified involves place of articulation. The different means used for analysing the data, thus, indicate to significance above chance for responses even thought the task involved a forced choice task. These indications based on descriptive statistics are further supported with the Multivariate Analyses of Variance (MANOVAs) that have preceded.

Table 1. Descriptive Statistics for Performance of Participants in terms of the Four Investigated Factors in Descending Order

	Rank Order	***M(SD)***	**N**
Variable 1: word-medial onset [p]	1	62.39(40.52)	113
Variable 2: word-initial onset [k]	2	60.47(38.84)	113
Variable 3: word-initial onset [g]	3	60.32(40.78)	113
Variable 4: word-initial onset [p]	4	59.82(42.47)	113

Variable 5: word-initial onset [t]	5	58.63(37.91)	113
Variable 6: word-final coda [k]	6	58.41(34.08)	113
Variable 7: word-medial onset [d]	7	56.93(35.13)	113
Variable 8: word-initial onset [d]	8	56.86(40.12)	113
Variable 9: word-medial onset [b]	9	56.19(39.46)	113
Variable 10: word-final coda [p]	10	54.72(35.18)	113
Variable 11: word-final coda [t]	11	54.65(37.58)	113
Variable 12: word-initial onset [b]	12	52.74(41.06)	113
Variable 13: word-medial onset [k]	12	52.74(42.18)	113
Variable 14: word-medial onset [g]	13	52.04(40.23)	113
Variable 15: word-medial onset [t]	14	51.03(35.79)	113
Variable 16: word-final coda [d]	15	38.50(38.53)	113
Variable 17: word-final coda [g]	16	36.58(34.42)	113
Variable 18: word-final coda [b]	17	34.81(33.45)	113

Discussion and concluding remarks

The results of the present study seem to straightforwardly be related to previous research in language acquisition. Specifically, a number of studies suggested that the less successful differentiation of plosive contrasts was the outcome of the investigated factors. L2 contrasts may often be difficult for users to perceive since they may not be skilled at attending to the needed acoustic cue or set of cues. Further, phonological contrasts in different L1s may be realised differently (acoustically speaking) that results in some degree of L1 influence on the weighting of acoustic cues in perception of the L2 users (depending on their L1 background). On the other hand, L1 users are able to slowly build perceptual categories by being exposed to meaningful input with no interference from an additional language. As a result, the examination of the different factors provides an indication of whether CG users are able to categorise L2 speech sounds into newly formed L2 phonological categories or whether they simply assimilate L2 speech sounds into existing L1 phonological categories. Based on the results, CG users must have perceived voicing but because the weighting of auditory cues in the categorisation of plosives is language-specific, they were modifying their identification of voiced plosives to fit the L1.

In this context, the acoustic cue of VOT seems to be of crucial importance when dealing with contrastive L2 categories (Okalidou et al., 2010; Tserdanelis & Arvaniti, 2001; Arvaniti, 1999; Lisker & Abramson, 1964). VOT, thus, provides important information for the voiceless-voiced distinction since voiceless plosives involve a longer VOT production compared to their voiced counterparts. Concerning place of articulation, velar plosives are produced with longer VOT values while bilabial are associated with the shortest VOT values. Next, plosives in onset position seem to be more easily perceived compared to coda position due to the acoustic cues referring to voicing and place of articulation that may not occur for plosives in coda position. Further, plosives in word initial position are not affected by phonological processes as in other positions.

Lastly, speech perception can also explain why the participants were more successful identifying velar consonants compared to their bilabial and alveolar counterparts since the former are associated with significantly longer VOT values than the other plosive consonants (Lisker & Abramson, 1964). Even if performance was also good concerning bilabial plosives, participants were generally more successful identifying velar plosives both voiceless and voiced while for bilabial plosives the same pattern was not observed since participants were better with voiceless bilabial plosives but not with their voiced counterparts. Since this study concentrated on plosive consonants in L2 English in a word-level, one issue that needs to be addressed in future research is whether a purely phonetic task involving only syllables such as an ABX test or a task involving words in sentences produce different results.

References:

Arvaniti, A. (2010). A (brief) overview of the phonetics and phonology of Cypriot Greek. *Proceedings of the conference on "The Greek language in Cyprus from antiquity to today, University of Athens*. August 2, 2017 Online: http://idiom.ucsd.edu/~arvaniti/ArvanitiCY.pdf

Arvaniti, A. (1999). Greek voiced stops: prosody, syllabification, underlying representations or selection of the optimal? *Proceedings of the 3rd International Linguistics Conference for the Greek Language*, Athens, Ellinika Grammata (pp. 383-390).

Botinis, A., Christofi, M., Themistocleous, C., & Kyprianou, A. (2004). Duration correlates of stop consonants in CG. *Proceedings, FONETIK 2004,* Dept. of Linguistics, Stockholm University.

Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, *20,* 384-422.

Okalidou, A., Petinou, K., Theodorou, E., & Karasimou, E. (2010). Development of voice onset time in standard-Greek and Cypriot-Greek-speaking preschoolers. *Clinical Linguistics & Phonetics*, *24*(7), 503-519.

Tserdanelis, G., & Arvaniti, A. (2001). The acoustic characteristics of geminate consonants in CG. *Proceedings of the Fourth International Conference on Greek Linguistics*, Thessaloniki, University Studio Press (pp.29-36).