Exploring Vocalization of /l/ in English: an EPG and EMA study

Mitsuhiro Nakamura

Nihon University, Tokyo, Japan

`nakamura.mitsuhiro99@nihon-u.ac.jp`

10.21437/Interspeech.2009-654

Abstract

This study explores the spatiotemporal characteristics of lingual gestures for the clear, dark, and vocalized allophones of /l/ in English by examining the EPG and EMA data from the multichannel articulatory (MOCHA) database. The results show the evidence that the spatiotemporal controls of the tip lowering and the dorsum backing gestures are organized systematically for the three variants. An exploratory description of the articulatory correlates for the /l/ gestures is made.

Index Terms: English /l/, clear/dark allophony, L-vocalization

1. Introduction

This study reports on the continuing investigation into the articulatory realization of the English /l/ and contributes to an understanding of the relationship between prosody and articulatory organization. It has been recognized that many varieties of English exhibit allophonic variation in /l/ [1]. The 'clear' variant occurs syllable-initially and the 'dark' variant occurs syllable-finally. In addition to the clear/dark allophony, the 'dark' allophone of /l/ can lose its consonantal nature, becoming a vowel of the [o] or [ʊ] type: this phenomenon is called L-vocalization. The current study focuses on the spatio-temporal properties of articulatory gestures for the vocalized /l/ relating them to those for the clear and dark variants of /l/.

Previous articulatory studies on the vocalized /l/ [2, 3, 4] have identified that the basic articulatory correlate is a withdrawal of the complete closure by the tip/blade of the tongue. Wrench & Scobbie [2] found that word-final /l/s show wide inter-speaker variation in vocalization rate and in the movement amplitude of the tongue tip/blade gesture. Hardcastle & Barry [3] pointed out that the dark/vocalized variation was non-categorical and sensitive to articulatory and perceptual factors. A hypothesis underlying these studies is that vocalization is not simply a coda weakening of the tip/blade gesture.

If dark /l/ and vocalized /l/ are closely related to each other as coda forms, not only the tip/blade but also the dorsum of the tongue should vary systematically between the clear, dark, and vocalized /l/. An influential study on the clear/dark allophony [5] has shown that the English /l/ can have more than one lingual constriction; that the dark allophone exhibits greater tongue dorsum retraction and lowering than the clear allophone; and that maximum dorsal movement precedes maximum apical movement for the dark variant, while the opposite temporal relationship occurs for the clear variant. This last characteristic of intergestural coordination is identified as 'tip delay.' It can be asked whether and how tip delay is realized when /l/ is vocalized.

The aim of this study is to further investigate the differential roles of the spatiotemporal articulatory parameters in the production of the clear, dark, and vocalized allophones in English /l/. Based on the EMA and EPG data from the multi-channel articulatory database, an exploratory description of the articulatory correlates for the three variants will be made.

2. Methodology

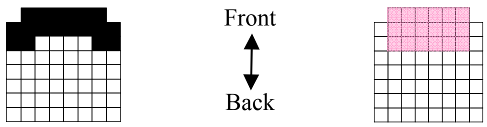
2.1. Data collection

The speech materials used in this study were drawn from the Multichannel Articulatory (MOCHA) database [6]. This database comprises articulatory and acoustic data of 460 phonetically-balanced sentences read by native speakers of English: each speaker read a set of 460 sentences once. Carstens Electromagnetic Articulograph (EMA, 500Hz sample rate), Laryngograph, and Reading Electropalatograph (EPG, 200Hz sample rate) were used for articulatory data acquisition. In the EMA data recording, the TT coil was placed at about 5-10mm, and the TD coil was placed at about 50-70mm from the tongue tip as measured with the tongue extended. The EPG and EMA data was analyzed for the selected utterances spoken by two speakers of Southern British English (RP) denoted below as SE (male) and SA (female).

Dataset in the current study consisted of monosyllabic words with the onset singleton /l/ (e.g. *l*ack, *l*arge, *l*ike), with the coda singleton /l/ (e.g. fai*l*, ski*ll*, te*ll*), and with the coda /lC/ cluster (e.g. be*lls*, bui*ld*, smi*les*). The nucleus vowel of these words was an unrounded monophthong or diphthong. When sampling the onset and the coda singleton /l/, a phonetic context was specified: the target item was preceded, or followed, by a schwa [ə] (e.g. 'a *l*arge,' 'the *l*ine' and 'ski*ll* of,' 'te*ll* an'). Other aspects of phonological contexts were not controlled. The MOCHA corpus was searched for examples which met the above requirements, giving 12 tokens for the onset singleton /l/, 8 tokens for the coda singleton /l/, and 30 tokens for the coda /lC/-cluster; in total, 50 tokens per speaker. Position in sentence of these tokens varied.

2.2. Articulatory analysis

The analysis of the EPG and EMA data was performed using a set of programs called EMA Tools [6]. The configurational characteristics were studied for EPG contact patterns in an alveolar zone. This zone, as shown in Figure 1 (a), covers the front three rows of 18 contacts. Following the index used by [4], alveolar contact in the production of /l/ was defined by reference to an '/l/-zone' in the three front rows on the palate as shown in Figure 1(b). If any contact occurred in this zone during the /l/ production, the token was classified as 'consonantal.' The EPG frame for the analysis was the one specified by the tangential velocity minimum of the EMA tongue tip trajectory. For some vocalized tokens, the minimum velocity of the dorsum trajectory was used as a reference point.



(a) Front (rows 1-3)
18/22=81.8%

(b) /l/-zone

Figure 1: *EPG prototypical palate and /l/ zone*

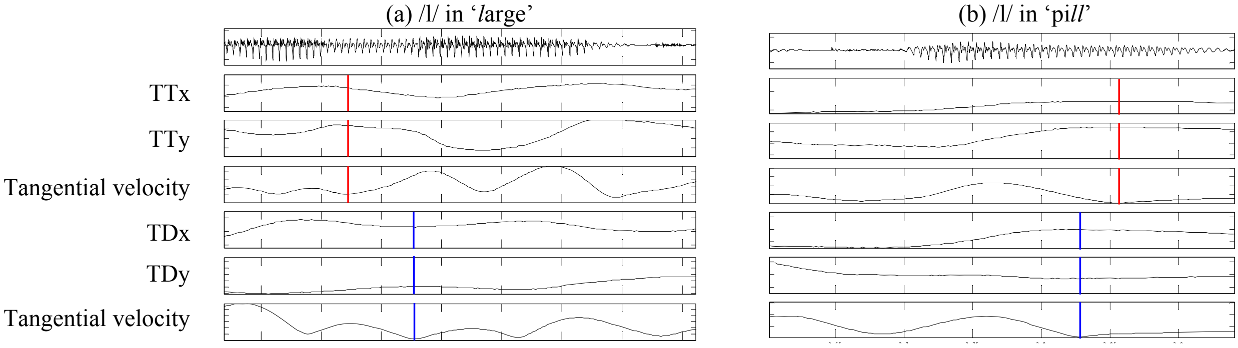

(a) /l/ in '*l*arge' (b) /l/ in 'pi*ll*'

Figure 2: *Examples of the landmark events identified in the EMA data analysis (speaker SE)*

Figure 2 shows the EMA recording of the tongue tip (TT) coil, the tongue dorsum (TD) coil, and the tangential velocity in the production of 'large' and 'pill.' Using the EMA Tools, time and position associated with an /l/ constriction were defined by the tangential velocity minimum of the TT coil and that of the TD coil (the red and blue line in Figure 2 respectively). For analyzing the spatial characteristics at the point of the articulators' minimum velocity, coordinate measurements in the horizontal and vertical dimensions were taken to specify the *constriction peak* of the tongue tip and dorsum: these are specified as the distance from the upper-incisor reference coil. For studying the temporal characteristics, the times of the minimum velocity of the TT and TD coils were used to calculate the timing relation, 'tip delay.' Following Sproat & Fujimura [5], *tip delay* in this study was defined as follows: [Tip Delay] = time (TT velocity minimum) − time (TD velocity minimum). Comparable TT and TD data was obtained from the majority of vocalized /l/ tokens which lack EPG contact. In cases with no clear TT minimum, the TD velocity minimum point was used as a reference.

Statistical comparisons were made by one- and two-way ANOVAS with Spyotvoll/Stoline post-hoc test ($p<0.05$). Only the main effects (speakers and three variants of /l/) were considered in this study. The statistical relationship between the variables was inferred from the results of the post-hoc test. The data collection was not controlled for the types of nuclear vowels, so their effects were not investigated.

3. Results

3.1. EPG analysis

In this analysis, vocalization and the configurational characteristics were examined. On the basis of the EPG data for all tokens, decisions were made as to whether the target /l/ was produced with alveolar contact (i.e. consonantal) or without it (i.e. vocalized), with reference to the /l/-zone index in Figure 1(b). The number of tokens for each subject is summarized in Table 1.

The distribution of consonantal and vocalized /l/ tokens shows that vocalization of the dark /l/ is typical of the word-final coda /lC/ clusters. They are composed of /l/ plus a dental [θ] (e.g. wea*l*th), one of the alveolars [t, d, s, z] (e.g. fe*l*t, spi*ll*ed, e*l*se, fi*l*es), one of the labials [p, f, m] (e.g. he*l*p, Ra*l*ph, fi*l*m), or a velar [k] (e.g. mi*l*k). In contrast, vocalization is less predictable in the word-final coda singleton /l/: only two instances were found for SA. This occurrence is compatible with the finding by [4]: word-final /l/ contact is onset-like before vowel-initial words.

Table 1. *Number of tokens for each speaker*

		Onset	Coda	
			singleton /l/	/lC/ cluster
SE	Consonantal	12	8	2
	Vocalized	0	0	28
SA	Consonantal	11	6	1
	Vocalized	1	2	29

For the purpose of comparison, the following EPG and EMA analyses were conducted for only the consonantal tokens in the onset and coda singleton /l/ groups and the vocalized tokens in the coda /lC/-cluster group. These groups are referred to as the onset /l/, the coda /l/, and the vocalized /l/ respectively in the rest of this study. Thus, the total number of tokens is 48 for SE and 46 for SA.

Figure 3 shows EPG contact patterns averaged separately for the onset, the coda, and the vocalized /l/ tokens. There are noticeable differences in the front region between the speakers: apical (SA) and apico-laminal (SE) realization. Figure 4 indicates the percentage values of contact in rows 1-3 for the onset /l/ and the coda /l/. The two-way ANOVA results demonstrate significant differences for the speakers [F(1,34)= 17.93, $p<0.01$] (SE>SA) and for the two variants [F(1,34)= 9.65, $p<0.01$] (onset /l/>coda /l/). The smaller amount of contact in the coda /l/ can be considered as syllable-final weakening. Also, there are differences in the amount of contact in the posterior region. This observation is compatible with that reported by an MRI study [8]: the tongue contact for the laminal /l/ extended laterally through the midpalatal region.

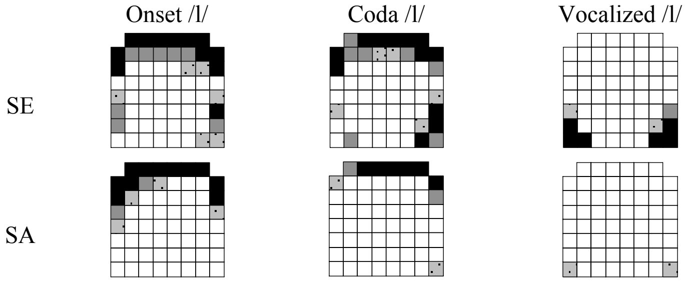


Figure 3: *EPG contact patterns*
black=100-80%; grey=79-50%; dotted=49-20%; blank=29-0%

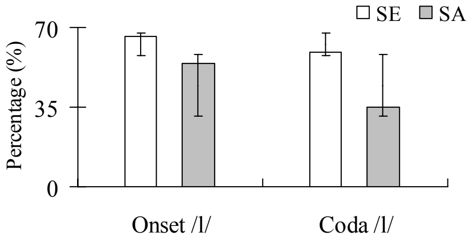


Figure 4: *Percentage of contact in the front region*
Error bars indicate ±1 standard deviation.

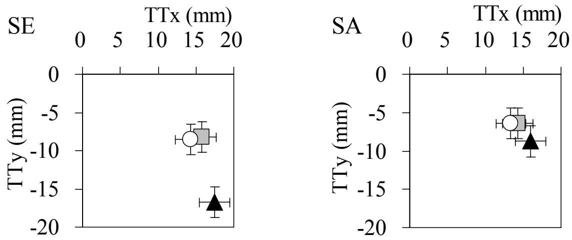


Figure 5: *Mean TTx/y displacement for /l/*
■=Onset /l/; ○=Coda /l/; ▲=Vocalized /l/
Error bars indicate ±1 standard deviation.

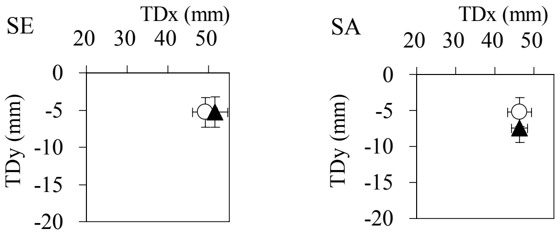


Figure 6: *Mean TDx/y displacement for /l/*
○=Coda /l/; ▲=Vocalized /l/
Error bars indicate ±1 standard deviation.

3.2. EMA analysis

In this analysis, we examine the spatiotemporal characteristics of the tongue tip and the dorsum gestures for the onset, coda, and vocalized /l/. The constriction peak of TTy and TDx and their relationship, and the temporal index for intergestural coordination, 'tip delay', are also investigated.

3.2.1. *Constriction peak of the tongue tip gesture*

Figure 5 above presents the mean TTx/y coordinates of constriction peak for the two speakers. While the positions are different for each speaker, they are not substantially divergent between the onset /l/ and coda /l/ within a speaker. Examination of the horizontal position (TTy) data shows the significant differences for the speakers [F(1,90)=75.22, p<0.01] and for the three /l/s [F(2,90)=34.82, p<0.01]; the post-hoc test shows that the vocalized /l/ exhibits the lowest horizontal position of the three: onset /l/=coda /l/>vocalized /l/.

A separate analysis of the TTy position was conducted for the onset and coda /l/ tokens on the one hand, and the vocalized /l/ tokens on the other. For the onset and coda tokens, a significant effect was found for the speakers [F(1,34) =36.97, p<0.01], SE(-8.3mm) < SA(-6.3 mm) but not for the two variants [F(1,34)=0.49, p=0.48], the onset /l/ (-7.3mm) = the coda /l/ (-7.6mm). For the vocalized /l/ tokens, a one-way ANOVA showed a significant difference between the speakers [F(1,55)=76.80, p<0.01]: the TTy position is markedly lower in SE(-16.77) than in SA(-8.74).

3.2.2. *Constriction peak of the tongue dorsum gesture*

Figure 6 above shows the mean TDx/y coordinates of constriction peak for the two speakers. Two characteristics are noticeable. One is that the TDx position, which is typically involved in the coda /l/ and vocalized /l/, is significantly different between the speakers [F(1,69)=85.42, p<0.01]: the TD peak is more posterior for SE (51.0mm) than for SA (46.2mm) in a pooled context of the two variants. The other characteristic is the changing direction of the dorsum position for the vocalized /l/. SE shows a more *posterior* position for the vocalized /l/ (51.5mm) than the coda /l/ (49.08mm), and SA shows a *lower* location for the vocalized /l/ (-7.4mm) than for the coda /l/ (-5.2mm). To examine the significance of these differences, a one-way ANOVA was conducted for the TDx data of SE and for the TDy data of SA. It was found that the peak positions were significantly different for SE [F(1,34)=6.79, p=0.013] but not for SA [F(1,33)=2.51, p=0.12]. This characteristic becomes more explicit if it is related to the tongue tip activities.

Figure 7 below plots the constriction peak of TTy relative to the constriction peak of TDx. For SE, the coda /l/ and vocalized /l/ tokens constitute separate groups, whereas for SA, the tokens overlap completely. A reasonable interpretation of these distributions is that there are different strategies for the vocalization of /l/, the choice of which depends on the speaker. Relative to the TT and TD gestures involved in the coda /l/, SA produces the vocalized /l/ primarily by lowering the tongue tip, whereas both tip lowering and dorsum backing are involved in SE's production of the vocalized /l/.

3.2.3. *Intergestural coordination: tip delay*

Figure 8 below reveals the mean values of tip delay for the onset, coda, and vocalized /l/ averaged for the two speakers. The general pattern of tip delay is not significantly different between the speakers [F(1,90)=0.002, p=0.95] but varies typically across the three variants [F(1,90)=20.38, p<0.01]. The value of tip delay is greater for the coda /l/ than for the onset /l/ and the vocalized /l/: the value pooled across the speakers increases in the order of onset /l/ (-35.0)<vocalized /l/ (-4.4)<coda /l/ (20.4). The negative value of the onset /l/ means that the constriction peak of the tongue dorsum is reached *after* that of the tongue tip. Conversely, the positive value of the coda /l/ means that the constriction peak of the tongue dorsum is reached *before* that of the tongue tip. Our results confirm those reported by [7]. It must be noted that there are idiosyncratic differences between the two coda forms. For SE, the timing pattern for the vocalized /l/ is distinctly

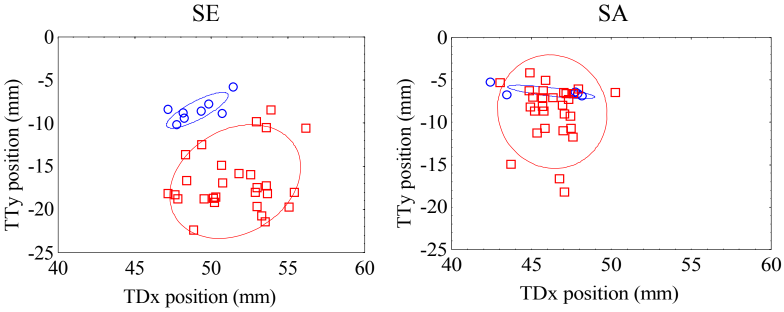


Figure 7: *TTy constriction peak positions plotted against TDx constriction peak position.* A 95% confidence region is drawn around the data for each variant:
○=Coda /l/; □=Vocalized /l/

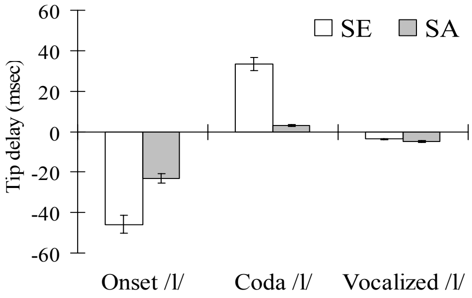


Figure 8: *Mean tip delay for the onset, coda, and vocalized /l/*
Error bars indicate ±1 standard deviation.

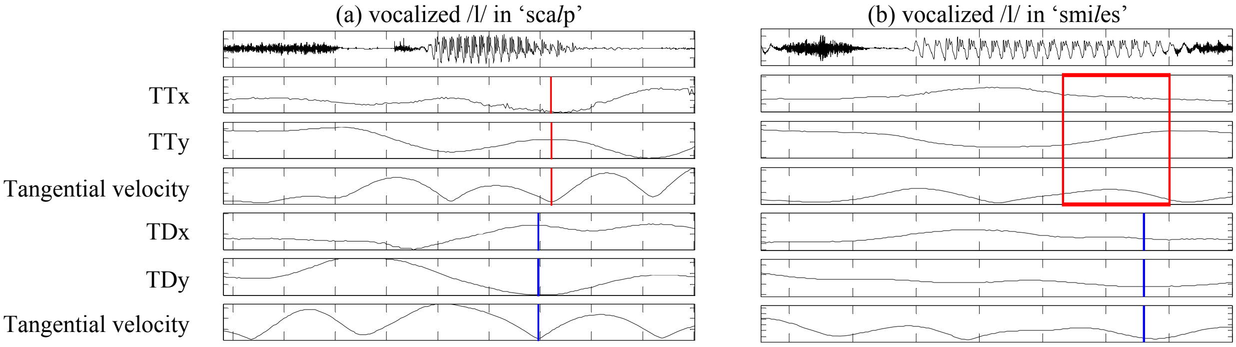

(a) vocalized /l/ in ‘sca*l*p’ (b) vocalized /l/ in ‘smi*l*es’

Figure 9: *Examples of the vocalized /l/ (speaker SA)*

different from that for the coda /l/. The negative value of the vocalized /l/ means that the constriction peak of the tongue tip is nearly synchronous with that of the dorsum. In contrast, such a significant tip-delay pattern was not observed for SA.

4. Discussion

The results presented so far are in general agreement with those obtained by previous articulatory studies on the clear/ dark allophony and vocalization of /l/ in English. Although our preliminary analysis is limited to lingual gestures and data for only two speakers, we have further evidence particularly from the vocalized /l/. The spatial variations in terms of the TT and TD constriction peaks showed that the speakers control the tip lowering and/or the dorsum backing as the production strategy for L-vocalization. The temporal variations in terms of tip delay revealed that the timing control of the tip and dorsum gesture constitutes a continuum for the onset, coda, and vocalized variants of /l/. Therefore, it is reasonable to suppose that the spatiotemporal properties of the tongue tip and dorsum gestures are organized, with reference to the syllable, to produce the qualitative differences in /l/.

It is certainly possible that speakers use other articulators or their combination with lingual gestures for vocalization of /l/. Although the speakers SE and SA showed different degrees of the TT lowering and the TD backing, the dorsum component always remained for the coda /l/ and vocalized /l/. Given that vocalization promotes the perceptual saliency of the coda (dark) /l/ [3] (i.e. a lower F2), the compressing or protruding gesture of the lips is one potential candidate. There is a difference of opinion about whether or not the labial gestures are involved in the dark /l/ and the vocalized /l/ [1, 2, 7]. A systematic examination of the roles of the lips would contribute to a better understanding of the two coda forms.

Another aspect of the tip lowering gesture and the dorsum backing gesture is concerned with coarticulation and phonology. Wells [1] mentioned that ‘…L Vocalization offers the prospect of eventual phonemic status for new diphthongs such as /ɪʊ/ (milk), /ɛʊ/ (shelf), etc’. We can point out some articulatory events observed in our vocalized /l/ tokens for further consideration. Given the tangential velocity minimum as an indicative of reaching an articulatory target, there are cases where the TT velocity trajectory shows no clear minimum point, indicating the transitional movement to the following consonant. As shown in Figure 9, the constriction peak of the tongue tip is attained in the vocalized /l/ in (a) ‘sca*l*p’ but not in (b) ‘smi*l*es’ (see a red box in Figure 9(b)): the tongue tip (TTy) is raised smoothly for preparing the following [z] while the tongue dorsum (TD) is reaching the target position for the vocalized /l/. Such a coarticulatory accommodation of the tongue tip was often observed in the coda /lC/ clusters, the final consonant of which is an alveolar. In addition to the loss of the consonantal component of the vocalized /l/, it is reasonable to assume that the tongue dorsum target position for the vocalized /l/, as specified by the velocity minimum, constitutes part of the preceding vowel or diphthong to form one large vocalic component. To understand these patterns fully, we probably need not only better phonetic models of coarticulation but also effective frameworks for analyzing diphthongs. Further research will provide useful implications for the relationship between L-vocalization and the vowel phonology of English.

5. Conclusion

This paper has investigated L-vocalization in terms of the spatiotemporal properties of lingual gestures. The preliminary results reveal that different properties for the onset, coda, and vocalized variants of /l/ are attributed to the systematic changes of the spatiotemporal parameters for the tip and the dorsum components of the tongue. The roles of the lips, the interaction between various kinematic parameters, and acoustic consequences will be substantiated by subsequent research.

6. Acknowledgements

I am very grateful to Prof. James Scobbie and Dr. Alan Wrench of Queen Margaret University College for their helpful suggestions for the use of the MOCHA database and EMA tools. This research is supported in part by a grant from Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research (C) 2009, Grant no. 21520415).

7. References

[1] Wells, J., Accents of English. 3 vols. Cambridge, 1982.
[2] Wrench, A.A. and Scobbie, J., “Categorizing Vocalization of English /l/ Using EPG, EMA, and ULTRASOUND”, Proc. the 6th International Seminar on Speech Production, Sydney, Australia, December 7 to 10, 314-319, 2003.
[3] Hardcastle, W. and Barry, W., “Articulatory and perceptual factors in /l/ vocalizations in English”, J. Int. Phon. Assoc., 15(2): 3-17, 1989.
[4] Scobbie, J.M., Pouplier, M., and Wrench, A., “Conditioning factors in external sandhi”, Proc. ICPhS XVI, Saarbrücken, Germany, 441-444, 2007.
[5] Sproat, R. and Fujimura, O., “Allophonic variation in English /l/ and its implications for phonetic implementation”, J. Phon., 21: 291-311, 1993.
[6] Wrench, A., “The MOCHA-TIMIT Articulatory Database”, Univ. Edinburgh, available at http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html., November, 1999.
[7] Giles, S.B. and Moll, K.L., “Cinefluorographic Study of Selected Allophones of English /l/”, Phonetica, 31: 206-227, 1975.
[8] Narayanan, S., Alwan, A.A., and Haker, K., “Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals”, J. Acoust. Soc. Am., 101(2): 1064-1077, 1997.