Durational Evidence for Syllable Boundary of /n/ and /l/ in Text-to-Speech Synthesis

Fang Tian
School of Cultures and Communications, Zhejiang University of Media and Communications, Hangzhou, China
E-mail: aconeywood@yahoo.com.cn

Abstract—**The Text-to-Speech (TTS) system does rely on syllable boundary information for segmental duration. However, ambisyllabic consonants always pose a problem to TTS because the system requires clear syllable boundaries to segment and concatenate. In order to provide a possible solution to this problem, */n/* and */l/* in $V_LC_AV_R$ are chosen in this paper as the target to be examined whether their durations behave more like the syllabic onset or coda when comparing with the durational properties of */n/* and */l/* both as onsets in *CV* and codas in *VC*. As the syllable boundaries, onset *C* shows much more sensitivity to stress than coda *C* while coda *C* shows more sensitivity to syllabic position than onset *C*. Moreover, C_A in $V_LC_AV_R$ is also influenced by two variables of stress and position as *C* in *CV* and *VC*. The results show that the intervocalic C_A holds the properties of both the syllabic onset and coda, which states the possibility that intervocalic consonants should be considered as a rather independent concatenative unit in TTS synthesis.**

Index Terms—**syllable, ambisyllabicity, $V_LN_AV_R$, $V_LL_AV_R$, duration, stress, position, onset, coda, text-to-speech synthesis**

I. Introduction

The correctness in pronunciation of unknown words is one of the keys to output more natural speech in the text-to-speech (TTS) system, one of whose elements do depend on syllable boundary for information, e.g. the segmental duration model. To produce high quality synthesis, the concatenation-based Text-to-Speech (TTS) system uses segments of recorded human [1], which usually demands a large stock of segmental units to cover various acoustic-phonetic contexts [2]. There is no doubt that the syllable structure plays a fundamental role in continuous speech [3]. For instance, by adopting syllable language models, a more precise phoneme recognition rate is expected in automatic speech recognition system [4][5][6][7]. Among all the rules to syllabify the English words, Syllable Onset Segmentation Hypothesis is the most outstanding rule which says syllable onsets are 'privileged alignment points' in word segmentation [8]. Syllable boundaries as both onsets and codas are, they are different acoustically, articulatorily and perceptually: For example, stop + vowel coarticulation across a syllable or word boundary (*C.V*) was found to be weaker than within a *CV* syllable, but were still greater than in *VC* in both English [9].

A. Ambisyllabicity and Contributing Factors

However, The syllable affiliation of intervocalic consonant clusters greatly exerts influences on V-to-V coarticulation [10], which directly leads to the phenomenon of ambisyllabicity, as indicated by common notations, such as CV[C]V, according to phonologists, that intervocalic consonant can be both syllabified as the coda of the previous syllable and the onset of the latter syllable, as /n/ in the word *pony*.

The syllabification of intervocalic consonants in English remains a controversial issue since phonologists favor different rules of syllabification, which arouses contradictions in one way or another. For instance, Selkirk [11] advocated an onset maximization strategy (as illustrated in /a.spar/) but Wells advocated a coda maximization strategy (as in /asp.ar/) [12]. In a word like *pony*, according to Charles W. Kreidler, the /n/ is ambisyllabic, simultaneously affiliated to both syllables. However, phonologists have different criteria for ambisyllabicity in even a single language such as English. Kahn places ambisyllabicity at the left edge of unstressed syllables [13]; Hooper places it at the right edge of stressed syllables [14].

As one way out of the predicament, many investigators have turned to psycholinguistic experiments for evidence of ambisyllabicity. Data from English, Danish and Dutch [15][16][17] indicate that ambisyllabicity may indeed be psychological real for speakers. In syllabifying tasks Treiman *et. al.* [18] found more ambisyllabic responses in words like *lemon* than *demon*, which contain short vowels in the first stressed syllable, and more ambisyllabic responses in words like *melon* than *wagon*, which contains sonorant consonants in the first stressed syllable.

As for the factors that may result in ambisyllabic responses, the stress pattern and vowel type are found to become the first two confounding factors, esp. a stressed syllable including a short vowel or an unstressed syllable including a diphthong or a long vowel, which are more likely to attract the intervocalic consonant to be their codas, thus giving rise to the phenomenon of ambisyllabicity. Besides, the consonant type is another factor. D. Fallows indicates that liquids and nasals, in particular, may be ambisyllabic in the minds of U.S. English [15]. Pharao N. discovered that the decisive factor of ambisyllabicity seems to be whether the consonant is a stop or a continuant [16], latter of which is more likely to be treated as ambisyllabic.

B. Experimental Variable

© 2013 ACADEMY PUBLISHER
doi:10.4304/jmm.8.2.82-89

Although the notion of ambisyllabicity exists in phonology and psycholinguistic, it’s rather helpless and impractical in speech engineering when precise syllabification is required. One possible solution is to find out the phonetic grounding for ambisyllabic consonant, that is, whether the prospective ambisyllabic segment actually has properties of the onset or coda, which is seldom reported in research. So in the present study, both as sonorants and continuants, the durational behaviors of the hypothetically ambisyllabic */n/* and */l/* in *VCV* sequences in speech corpus will be compared with durational behaviors of */n/* and */l/* as syllable boundaries, namely, both consonants in *NV*, *LV* and *VN*, *VL* sequences, to find out whether their durations behave phonetically more like the syllable onset or coda. Furthermore, the consonant duration is also influenced by the position of its belonging syllable in the multi-syllabic word. Takasawa, M. found in Spanish that nasal and liquid codas are longer than their onsets [19].

The speech corpus used in this experiment is naturally meaningful sentences, good at maintaining the naturalness, with the hope that the authentic results may be practical in future speech engineering. But with the disadvantage of confounding the influencing factors [20], all these factors in the corpus should be carefully considered as variables to find out their influences on the durations of */n/* and */l/* respectively. These factors basically are the stress pattern, the syllabic position and the vowel type according to the previous findings.

As for the last variable of vowel type, the number of *NV, VN, VNV* sequences with the five monophthongs /ɑ/, /æ/, /ʌ/, /ɪ/ and /ə/ in are grossly calculated and shown in Table 1.

TABLE 1. THE GROSS CALCULATION OF *NV, VN, VNV* WITH THE FIVE VOWELS

VNV	*VN*	count	*VNV*	*NV*	count
/ɑː$^{0/1}$n ə/	/ɑː $^{0/1}$ n /	416	/ə n ɑː$^{0/1}$ /	/ə^{0} n/	1015
	/nə^{0} /	5		/n ɑː $^{0/1}$ /	244
/æ$^{0/1}$n ə/	/æ$^{0/1}$ n/	296	/ə n æ$^{0/1}$/	/ə^{0} n/	1015
	/nə^{0} /	5		/n æ$^{0/1}$/	25
/ʌ$^{0/1}$ n ɪ/	/ʌ$^{0/1}$n/	66	/ɪ n ʌ$^{0/1}$/	/ɪ n/	76
	/n ɪ/	17		/n ʌ$^{0/1}$/	56
/æ$^{0/1}$ n ɪ/	/æ$^{0/1}$ n/	296	/ɪ n æ$^{0/1}$/	/ɪ n/	76
	/n ɪ/	17		/n æ$_{1/2}$/	25
/ʌ$^{0/1}$n iː/	/ʌ$^{0/1}$n/	66	/ə^{0} nə^{0} /	/ə^{0} n/	1015
	/n iː/	119		/nə^{0} /	5

0--Unstressed; 1--stressed

The five vowels are found to distribute quite unevenly, which means a strict control of the vowel type won’t guarantee enough quantity in the experiment. In fact, the influence on consonant duration by vowels will be balanced out if all vowel types appear almost evenly over the target phonetic sequences. So the vowel type will be a less important variable to be discussed in this experiment.

C. *Experimental Hypothesis*

Certain rules are set for the variables. Firstly, there are only two stress levels, stressed or unstressed. The secondary stress is still considered as stressed. Secondly, the relative positions of *CV* and *VC* sequences in multi-syllabic words are classified into the initial, the middle and the final positions in multisyllabic words. Thirdly, positions of *VNV* and *VLV* will be roughly classified into four kinds: (1) V_L^0 in the initial position of a multi-syllabic word while V_R^1 in the middle position; (2) V_L^0 in the initial position while V_R^1 in the final position; (3) both V_L^0 and V_R^1 in the middle positions; (4) V_L^0 in the middle position while V_R^1 in the final position. Among all the symbols, the superscripts 0 and 1 represent the stress level with 0 as unstressed and 1 as stressed; the subscripts *L* and *R* represent the consonant position in the syllable with *L* as onset and *R* as coda.; subscript *A* represent the hypothetically ambisyllabic consonant.

Finally, the research hypothesis is proposed: if *C* in V_LCV_R is more closely bound to V_R, it should durationally behave more like the syllabic onset, as *C* in *CV*; if *C* is more closely bound to V_L, it should durationally behave more like the syllabic coda, as *C* in *VC*.

II. Experimental Results

A. *Corpus*

In this experiment, the natural connected speech from the MSRA speech corpus is used. It includes 6,000 meaningful and connected sentences from a native American female.

B. *Experimental Process*

A program is applied to pick out all the legal phonetic sequences from it. All the durational results are first tagged by SFS software and then refined by a phonetician while listening to the original sound and watching the corresponding spectrograms. Finally, all the quantity of *CV, VC* and *VCV* sequences are counted out with legal 1110 */n/* sequences and 1183 */l/* sequences. And the durations for each type are calculated out and exhibited in the EXCEL. The following are some example spectrograms of */n/* and */l/* in *CV, VC* and *VCV* sequences from the qualified sequences shown in fig. 1-4.

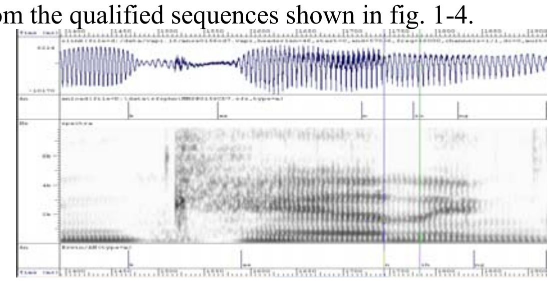

Fig. 1 /n/ in *canning* from *buying canning supplies*

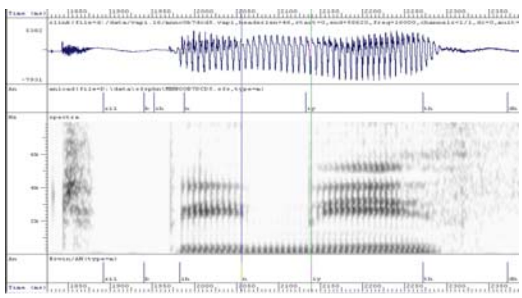

Fig. 2 /n/ in *beneath* from *powder flame gushed beneath the wagon*

© 2013 ACADEMY PUBLISHER

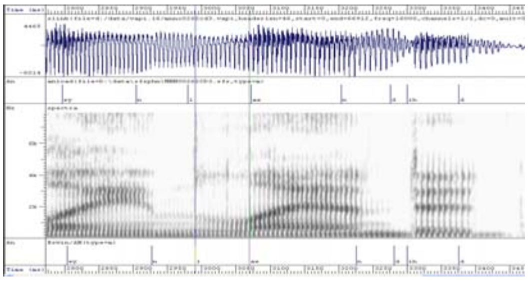

Fig. 3 The light /l/ in *landed* from *after a long trip the airplane landed*

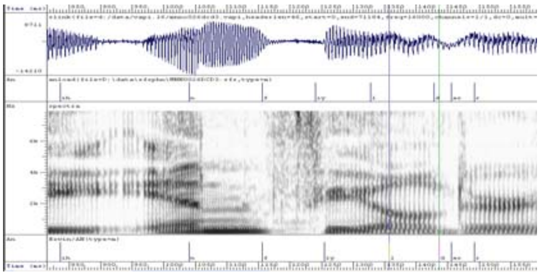

Fig. 4 The dark /l/ in *infielder* from *infielders dislike their baseline hits*

III. DISCUSSION

All the counts of legal sequences are illustrated in TABLE 2. The analytical process includes two steps: 1) the durational properties of syllable onsets and codas of /n/ and /l/ will be summarized first as the comparative group; 2) the durational properties of hypothetically ambisyllabic */n/* and */l/* in *VCV* sequence will be compared with the durational properties of */n/* and */l/* as onsets and codas for a further comparative study.

TABLE 2. THE SYLLABLE COUNT FOR *CV, VC* AND *VCV* SEQUENCES

C	*VNV*	count	*CV*	count	*VC*	count
/n/	$V_L^0\ N_A\ V_R^0$	125	$N_L^0 V_R^0$	40	$V_L^0\ N_R^0$	271
	$V_L^0\ N_A\ V_R^1$	231	$N_L^0 V_R^1$	154	$V_L^1\ N_R^0$	240
	$V_L^1\ N_A\ V_R^0$	49				
/l/	$V_L^0\ L_A\ V_R^0$	34	$L_L^0\ V_R^0$	5	$V_L^0\ L_R^0$	154
	$V_L^0\ L_A\ V_R^1$	324	$L_L^0\ V_R^1$	225	$V_R^1\ L_R^0$	70
	$V_L^1\ L_A\ V_R^0$	371				

count—numbers of the qualified phoneme in the sequence; ^{0}—unstressed; ^{1}—stressed; $_L$—left position in the *VCV* sequence; $_R$—right position in the *VCV* sequence; $_A$—the hypothetically ambisyllabic consonant

A. *Durational Properties of /n/ and /l/ as Onsets and Codas*

1) Ddurational analysis of onset/n/ in NV and coda /n/ in VN

Table 3 demonstrates the count, mean duration (msec) and SD for /n/ as onsets and codas with different stress levels in the initial, middle and final positions respectively.

TABLE 3. THE COUNT, MEAN DURATION AND SD OF */n/* WITH DIFFERENT STRESS AND POSITION

N_L & N_R	N_L^0	N_R^0	N_L^1	N_R^1
count	40	271	154	240
mean	53.89	67.95	75.47	62.41
SD	19.90	26.43	21.11	27.68
N_L & N_R -i	N_L^0 -i	N_R^0 -i	N_L^1 -i	N_R^1 -i
count	22	48	104	90
mean	55.07	79.76	76.81	61.03
SD	18.57	18.43	20.15	19.62
N_L & N_R -m	N_L^0 -m	N_R^0 -m	N_L^1 -m	N_R^1 -m
count	16	107	42	62
mean	50.64	48.80	73.87	43.96
SD	22.50	16.38	24.62	15.36
N_L & N_R -f	N_L^0 -f	N_R^0 -f	N_L^1 -f	N_R^1 -f
count	2	116	8	88
mean	66.9	77.59	66.37	76.83
SD	8.50	24.00	9.09	33.03

-i--the initial position in the multi-syllabic word; -m--the middle position; -f--the final position **count**—numbers of the qualified phoneme in the sequence; **mean**—the mean duration of the phoneme; **SD—**the standard derivation of the duration results

First, the durations of onset */n/* are found to be almost the same in three positions with the general mean value. Then one-way ANOVA is adopted to check the influence of stress and position on the onset */n/* as shown in Table 4 and TABLE 5.

TABLE 4. THE ONE-WAY ANOVA AYALSIS OF POSITION AS THE VARIABLE IN *NV* WITH THE SAME STRESS LEVEL

N_L^0 & N_L^1	**Initial** *vs* **Middle**	**Initial** *vs* **Final**	**Middle** *vs* **Final**
N_L^0	$F_{0.05}$ (1,36)=0.44 (p<1)	$F_{0.05}$ (1,22)=0.77 (p<1)	$F_{0.05}$ (1,16)=0.98 (p<1)
N_L^1	$F_{0.05}$ (1,144)=0.56 (p<1)	$F_{0.05}$ (1,110)=2.1 (p<1)	$F_{0.05}$ (1,48)=0.71 (p<1)

Initial--the initial position in the multi-syllabic word; **Middle**--the middle position; **Final**--the final position

TABLE 5 THE ONE-WAY ANALYSIS OF STRESS AS THE VARIABLE IN *NV* AT THE SAME POSITION

N_L^0 & N_L^1	**General**	**Initial**	**Medial**
F	$F_{0.05}$ (1,192)=33.94 (p<0.001)	$F_{0.05}$ (1,124)=21.7 (p<0.001)	$F_{0.05}$ (1,56)=21.29 (p<0.001)

General--with no variable considered

Except the fact that N_R^0 in the initial position and N_R^1 in the final position show no significant difference in duration in Table III, the remaining results all point out that onset /n/ is insensitive to position with the same stress level. By tracing back the original words which extract N_R^0, we can easily find out this category is completely composed of compound adjective with prefix *un* before them, such as *unfit*, *uncover*. */n/* under this context may, more properly, be considered as the word

© 2013 ACADEMY PUBLISHER

coda, rather than the syllable coda in the speaker's mind, which makes it considerably longer. In contrast, with the same position, the mean duration is remarkably longer when */n/* is stressed in Table 4. That is to say, onset */n/* is sensitive to stress with the same position.

TABLE 6 THE ONE-WAY ANALYSIS OF POSITION AS THE VARIABLE IN *VN* WITH THE SAME STRESS LEVEL

N_R^0 & N_R^1	Initial *vs* Middle	Initial *vs* Final	Middle *vs* Final
N_R^0	$F_{0.05}$ (1,153)=109.43 (p<0.001)	$F_{0.05}$ (1,162)=0.32 (p<1)	$F_{0.05}$ (1,221)=107.7 (p<0.001)
N_R^1	$F_{0.05}$ (1,150)=33.02 (p<0.001)	$F_{0.05}$ (1,176)=15.13 (p<0.001)	$F_{0.05}$ (1,148)=53.23 (p<0.001)

Differently, the mean durations of coda */n/* in general and three positions are various with the shortest in the middle position and rather longer in both initial and final positions. In TABLE 6, the one-way ANOVA proves this result, which is to say that coda */n/* is sensitive to position with the same stress level.

However, with the same position, the mean duration for coda */n/* is not remarkably longer when */n/* obtains the stress. In TABLE 7, the one-way ANOVA proves this result, which is to say that coda */n/*'s sensitivity to stress is not as sharp as onset */n/* with the same position.

TABLE 7 THE ONE-WAY ANALYSIS OF POSITION AS THE VARIABLE IN *VN* WITH AT THE SAME POSITION

N_R^0 & N_R^1	General	Initial	Medial	Final
F	$F_{0.05}$ (1,509)=3.26 (p<0.01)	$F_{0.05}$ (1,136)=29.74 (p<1)	$F_{0.05}$ (1,167)=3.59 (p<0.1)	$F_{0.05}$ (1,202)=0.04 (p<1)

2) *The durational analysis of onset /l/ in LV and coda/l/ in VL*

The count, mean duration (msec) and SD for */l/* as onsets and codas with different stress levels in three positions respectively are shown in Table 8.

TABLE 8. THE COUNT, MEAN DURATION AND SD OF /l/ WITH DIFFERENT STRESS AND POSITION

L_L & L_R	L_L^0	L_R^0	L_L^1	L_R^1
count	59	154	225	70
mean	48.51	80.15	69.97	92.64
SD	18.08	34.90	26.13	34.09
L_L & L_R -i	L_L^0 -i	L_R^0 -i	L_L^1 -i	L_R^1 -i
count	5	44	160	17
mean	68.44	74.68	81.53	83.38
SD	15.18	20.09	18.59	16.98
L_L & L_R -m	L_L^0 -m	L_R^0 -m	L_L^1 -m	L_R^1 -m
count	18	24	58	17
mean	30.10	56.98	39.60	70.22
SD	13.72	15.64	18.38	17.64
L_L & L_R -f	L_L^0 -f	L_R^0 -f	L_L^1 -f	L_R^1 -f
count	36	86	7	36
mean	54.94	89.42	57.23	107.60
SD	12.27	40.842	21.382	38.74

TABLE 9. THE ONE-WAY ANOVA AYALSIS OF POSITION AS THE VARIABLE IN *IV* WITH THE SAME STRESS LEVEL

L_L^0 & L_L^1	Initial *vs* Middle	Middle *vs* Final
L_L^0	None	$F_{0.05}$ (1,52)=26.76(p<0.001)
L_L^1	$F_{0.05}$ (1,216)=217.92(p<0.001)	None

With the same stress pattern, onset */l/* is still sensitive to position with the longest duration in the initial position. This is especially true when *LV* is the real first syllable of a multi-syllabic word as in *loudly* or *luxurious*. Although the one-way ANOVA cannot check all the three positions because of the small number for L_L^0-i and L_L^1-f, the results in TABLE 9 show that onset */l/* is generally sensitive to position with the same stress level.

Moreover, stress is generally the other variable to influence its duration, which is proved by the one-way ANOVA in TABLE 10.

TABLE 10 THE ONE-WAY ANALYSIS OF STRESS AS THE VARIABLE IN *LV*

L_L^0 & L_L^1	General
F	$F_{0.05}$ (1,211)=132.29(p<0.001)

Likewise, the duration of coda */l/* is sensitive to position with the longest in the initial position and shortest in the middle position. In TABLE 11, the one-way ANOVA proves this result, which is to say that coda */l/* is almost sensitive to position with the same stress level.

TABLE 11 THE ONE-WAY ANALYSIS OF POSITION AS THE VARIABLE IN *VL* WITH THE SAME STRESS LEVEL

L_R^0 & L_R^1	Initial *vs* Middle	Initial *vs* Final	Middle *vs* Final
L_R^0	$F_{0.05}$ (1,66)=13.97 (p<0.001)	$F_{0.05}$ (1,128)=5.09 (p<0.05)	$F_{0.05}$ (1,108)=14.17 (p<0.001)

With the same position, coda */l/* lengthens when it is stressed. However, the result of one-way ANOVA in TABLE 12 shows that sensitivity of coda */l/* decreases comparing with onset*/l/*.

TABLE 12 THE ONE-WAY ANALYSIS OF STRESS AS THE VARIABLE IN *VL* AT THE SAME POSITION

L_R^0 & L_R^1	General
F	$F_{0.05}$ (1,222)=6.25(p<0.02)

In conclusion, both as sonorants, */n/* and */l/* show different durational features. Onset */n/* is sensitive to stress, but not position; coda */n/* is sensitive to position, not so much to stress. Overall, */n/* shows its relative stability as both the syllable onset and coda. By contrast,

© 2013 ACADEMY PUBLISHER

/l/ exhibits less stability in the fact that different syllabic position influences its duration. The duration of /l/ is greatly influenced by the neighboring vowel no matter it's the onset or the coda. But generally speaking, onset /l/ is found to be much more sensitive to stress than coda /l/.

3) *The durational analysis of V in CV and VC*

Although the vowel type isn't controlled in this experiment and therefore cannot be considered as a reliable variable in analysis, its durational changes will reflect its general tendency under different stress levels and positions, which are shown in TABLE 13 and TABLE 14.

TABLE 13 THE COUNT, MEAN DURATION AND SD OF *V* IN *NV* AND *VN* WITH DIFFERENT STRESS AND POSITION

$_{N}V_{L}$ & $_{N}V_{R}$	V_R^0	V_L^0	V_R^1	V_L^1
count	40	271	154	240
mean	87.22	73.10	118.20	124.53
SD	47.67	38.93	38.25	47.27
V_L **&** V_R **-i**	V_R^0 **-i**	V_L^0 **-i**	V_R^1 **-i**	V_L^1 **-i**
count	22	48	104	90
mean	106.01	107.00	123.54	107.17
SD	46.24	19.49	38.34	40.20
V_L **&** V_R **-m**	V_R^0 **-m**	V_L^0 **-m**	V_R^1 **-m**	V_L^1 **-m**
count	16	107	42	62
mean	55.46	48.00	102.00	100.57
SD	27.95	15.61	36.39	18.83
V_L **&** V_R **-f**	V_R^0 **-f**	V_L^0 **-f**	V_R^1 **-f**	V_L^1 **-f**
count	2	116	8	88
mean	134.61	82.22	133.82	159.16
SD	59.97	45.05	17.93	47.98
V_L **&** V_R **-f**	V_R^0 -f	V_L^0 -f	V_R^1 -f	V_L^1 -f
count	36	86	7	36
mean	136.16	59.68	181.98	112.18
SD	79.02	45.05	63.61	44.317

TABLE 14 THE COUNT, MEAN DURATION AND SD OF *V* IN *LV* AND *VL* WITH DIFFERENT STRESS AND POSITION

$_{L}V_{L}$ & $_{L}V_{R}$	V_R^0	V_L^0	V_R^1	V_L^1
count	59	154	225	70
mean	105.15	66.64	112.39	94.50
SD	73.75	42.34	40.75	40.60
V_L **&** V_R **-i**	V_R^0 **-i**	V_L^0 **-i**	V_R^1 **-i**	V_L^1 **-i**
count	5	44	160	17
mean	61.26	98.17	116.24	87.77
SD	15.94	23.13	37.35	23.35
V_L **&** V_R **-m**	V_R^0 **-m**	V_L^0 **-m**	V_R^1 **-m**	V_L^1 **-m**
count	18	24	58	17
Mean	55.33	33.77	93.38	63.80
SD	20.21	14.67	34.68	22.67

V_L **&** V_R **-f**	V_R^0 **-f**	V_L^0 **-f**	V_R^1 **-f**	V_L^1 **-f**
Count	36	86	7	36
Mean	136.16	59.68	181.98	112.18
SD	79.02	45.05	63.61	44.317

In the following tables, the duration changes accordingly when both the stress and position change. This trend is universal whenever /n/ and /l/ occupy the onset or coda positions. That is to say, vowel duration is sensitive to both stress level and syllabic position.

B. *Durational Tendency of Hypothetical C in VCV Sequences*

1) *Durational analysis of N_A in $V_LN_AV_R$*

The count, mean duration (msec) and SD of $V_LN_AV_R$ with different stress levels and positions are presented respectively in the following TABLE 14, 15, 16 and 17 .

TABLE 14. V_L^0 IN THE INITIAL AND V_R^1 IN THE MIDDLE

$V_L^0\ N_A\ V_R^1$	**count**	**mean**	**SD**
V_L^0	105	70.86	26.47
N_A	105	**70.18**	13.08
V_R^1	105	123.67	31.93
$V_L^1\ N_A\ V_R^0$	**count**	**mean**	**SD**
V_L^1	150	102.01	31.63
N_A	150	**39.08**	10.71
V_R^0	150	54.25	24.32

TABLE 15 V_L^0 IN THE INITIAL AND V_R^1 IN THE FINAL

$V_L^0\ N_A\ V_R^1$	**count**	**mean**	**SD**
V_L^0	41	63.68	23.74
N_A	41	**67.10**	11.41
V_R^1	41	163.45	68.2
$V_L^1\ N_A\ V_R^0$	**count**	**mean**	**SD**
V_L^1	213	111.12	39.86
N_A	213	**44.33**	9.11
V_R^0	213	84.13	36.21

TABLE 16 BOTH V_L^0 AND V_R^1 IN THE MIDDLE

$V_L^0\ N_A\ V_R^0$	**count**	**mean**	**SD**
V_L^0	19	33.63	22.00
N_A	19	**41.03**	11.31
V_R^0	19	73.05	27.48
$V_L^0\ N_A\ V_R^1$	**count**	**mean**	**SD**
V_L^0	57	44.97	20.37
N_A	57	**55.93**	10.13
V_R^1	57	111.48	24.82
$V_L^1\ N_A\ V_R^0$	**count**	**mean**	**SD**
V_L^1	44	103.22	36.57
N_A	44	**43.11**	11.72

© 2013 ACADEMY PUBLISHER

V_R^0	44	49.29	20.71

TABLE 17 BOTH V_L^0 AND V_R^1 IN THE MIDDLE

$V_L^0\ N_A\ V_R^0$	count	mean	SD
V_L^0	101	41.44	18.07
N_A	101	**48.08**	11.20
V_R^0	101	91.94	65.97
$V_L^0\ N_A\ V_R^1$	count	mean	SD
V_L^0	34	44.76	12.13
N_A	34	**56.21**	12.89
V_R^1	34	154.32	113.20
$V_L^1\ N_A\ V_R^0$	count	mean	SD
V_L^1	49	126.67	33.54
N_A	49	**43.76**	8.17
V_R^0	49	78.75	34.84

In the following analysis, the durational properties of N_A, V_L and V_R depending on stress and position are checked by one-way ANOVA, and then compared with the durational properties of */n/* in *NV* and *VN*.

First, an obvious interaction holds between N_A's duration and $V_LN_AV_R$'s position, with the shortest when both vowels are in the middle position and longer in the other three conditions. Second, when there are only two stress patterns in TABLE 14&15, similar durational changes of N_A are found in that N_As are significantly longer when only V_Rs are stressed, which are proved by the one-way ANOVA respectively ($F(1, 253)=433.17$ ($p<0.001$); $F(1, 252) = 196.91$ ($p<0.001$)). By adding the $V_L^0\ N_A\ V_R^0$ in TABLE 16&17, different durational tendency of N_A is found in that N_As show no significant durational changes when both V_L's stress patterns change from $V_L^0\ N_A\ V_R^0$ to $V_L^1\ N_A\ V_R^0$. Then one-way ANOVA in TABLE 18 proves the findings.

TABLE 18 THE ONE-WAY ANOVA OF N_A IN $V_LN_AV_R$

VNV Pair	Position	F
$V_L^0\ N_A\ V_R^0 — V_L^0\ N_A\ V_R^1$	M—M	$F(1,74)=29.08$($p<0.001$)
	M—F	$F(1,133)=12.40$($p<0.001$)
$V_L^0\ N_A\ V_R^0 — V_L^1\ N_A\ V_R^0$	M—M	$F(1,61)=0.43$($p<1$)
	M—F	$F(1,148)=5.78$($p<0.02$)
$V_L^0\ N_A\ V_R^1 — V_L^1\ N_A\ V_R^0$	M—M	$F(1,99)=34.66$($p<0.001$)
	M—F	$F(1,81)=28.97$($p<0.001$)

M-- middle; F--final

Compared with the durational properties of */n/* in *NV* and *VN*, N_A in $V_LN_AV_R$ shows a two-way tendency in duration. On one hand, by being sensitive to the stress pattern of V_R, it seems more closely bound to V_R, which behaves partially the same with onset */n/* in *NV*. On the other hand, in $V_L^0\ N_A\ V_R^0$ and $V_L^1\ N_A\ V_R^0$, N_A shows no significant changes whatever stress pattern V_L is, which happens partially to be the durational properties of coda */n/* in *VN*.

2) Durational Analysis of LA in VLLAVR

The count, mean duration (msec) and SD of $V_LL_AV_R$ are presented in the same way with *VNV*, as are illustrated in TABLE VII, VIII, IX and X.

TABLE 19 V_L^0 IN THE INITIAL AND V_R^1 IN THE MIDDLE

$V_L^0\ L_A\ V_R^1$	count	mean	SD
V_L^0	102	59.73	20.19
L_A	102	**76.65**	16.96
V_R^1	102	93.85	25.87
$V_L^1\ L_A\ V_R^0$	count	mean	SD
V_L^1	80	104.87	32.41
L_A	80	**51.90**	10.81
V_R^0	80	60.22	19.18

TABLE 20 V_L^0 IN THE INITIAL AND V_R^1 IN THE FINAL

$V_L^0\ L_A\ V_R^1$	count	mean	SD
V_L^0	90	62.80	17.62
L_A	90	**78.11**	14.95
V_R^1	90	180.44	50.14
$V_L^1\ L_A\ V_R^0$	count	mean	SD
V_L^1	182	130.31	38.41
L_A	182	**57.45**	12.88
V_R^0	182	80.42	63.74

TABLE 21 BOTH V_L^0 AND V_R^1 IN THE MIDDLE

$V_L^0\ L_A\ V_R^1$	count	mean	SD
V_L^0	85	44.69	17.49
L_A	85	**59.91**	15.68
V_R^1	85	109.65	31.21
$V_L^1\ L_A\ V_R^0$	count	mean	SD
V_L^1	73	113.83	28.21
L_A	73	**51.29**	10.37
V_R^0	73	59.82	15.56

TABLE 22 V_L^0 IN THE MIDDLE AND V_R^1 IN THE FINAL

$V_L^0\ L_A\ V_R^0$	count	mean	SD
V_L^0	30	42.39	15.98
L_A	30	**59.19**	16.41
V_R^0	30	106.79	52.68
$V_L^0\ L_A\ V_R^1$	count	mean	SD
V_L^0	47	50.18	14.80
L_A	47	**69.84**	15.82
V_R^1	47	173.55	108.81
$V_L^1\ L_A\ V_R^0$	count	mean	SD
V_L^1	36	111.65	33.80
L_A	36	**60.89**	12.63

© 2013 ACADEMY PUBLISHER

V_R^0	36	73.29	19.23

L_A shows similar durational changes as N_A. First, an obvious interaction holds between L_A's duration and $V_L L_A V_R$'s position, with the shortest when both vowels are in the middle position and longer in the other three conditions. Second, when there are only two stress patterns in TABLE 19&20, similar durational changes of L_A are found in that N_As are significantly longer when only V_Rs are stressed, which are proved by the one-way ANOVA respectively (F(1,180)=129.09(p<0.001); F(1,270)=139.07(p<0.001)).

By adding the $V_L^0\ L_A\ V_R^0$ in TABLE 21&22, different durational tendency of L_A is found in that L_As show no significant changes when both V_L's stress patterns change from $V_L^0\ L_A\ V_R^0$ to $V_L^1\ L_A\ V_R^0$. Then one-way ANOVA in TABLE 23 proves the findings.

Compared with the durational properties of */l/* in *LV* and *VL*, L_A in $V_L L_A V_R$ also shows a two-way tendency in duration. On one hand, by being sensitive to the stress pattern of V_R, L_A seems more closely bound to V_R, which behaves partially the same with onset */l/* in *LV*. On the other hand, in $V_L^0\ L_A\ V_R^0$ and $V_L^1\ L_A\ V_R^0$, L_A shows no significant changes whatever stress pattern V_L is, which happens partially to be the durational properties of coda */l/* in *VL*.

TABLE 23 THE ONE-WAY ANOVA OF L_A IN $V_L L_A V_R$

VLV	Position	F
$V_L^0\ L_A\ V_R^0 - V_L^0\ L_A\ V_R^1$	M—-F	$F(1,75)=8.06(p<0.01)$
$V_L^0\ L_A\ V_R^0 - V_L^1\ L_A\ V_R^0$	M—-F	$F(1,64)=0.23(p<1)$
$V_L^0\ L_A\ V_R^1 - V_L^1\ L_A\ V_R^0$	M—-M	$F(1,156)=16.05(p<0.001)$
	M—-F	$F(1,81)=7.73(p<0.01)$

VI. Conclusion

Based on the durational behaviors of */n/* and */l/* as the syllable onset or coda, this paper attempts to find the durational properties of intervocalic N_A and L_A, hoping to discover the syllabic boundary, thus providing the evidence for segmentation.

Although both are sonorants, they behave differently in *CV* and *VC* sequences. On one hand, onset */n/* is sensitive to the stress, but not the position while coda */n/* is sensitive to the position but not so much to the stress. Therefore, in speech synthesis, the constraint of stress will be set to select onset */n/*, but not the position while the constraint of position will be set to select coda */n/*, but not the stress. On the other hand, */l/* bears more variations than */n/* under the same condition. For */l/*, both onset and coda are influenced by stress and position. That is to say, */l/* is found to be more complicated and unstable than */n/* as both onsets and codas. In both cases, the vowel duration is sensitive to both stress level and syllabic position, which proves that the vowel under different contexts should be considered in producing more natural speech in TTS.

It is precise to say that stress and position cast certain effects on the durational tendency of N_A and L_A in $V_L C_A V_R$ in the same manner. Both consonants own the shortest duration when V_L and V_R are in middle position and longer ones in the other conditions. More importantly, C_A lengthens when V_R is stressed, whose behaviors are more like syllabic onset, meanwhile shows no significant durational changes when V_L is stressed, which accords more with syllabic coda. This implies the double-role of C_A at the very moment, holding both the properties of syllabic onset and coda, which proves the phenomenon of ambisyllabicity. As Content, A. *et.al.* pointed out, French resyllabified consonants in *VCV* sequences adopt the durational properties of neither syllabic onsets or word-initial consonants, but maintain the durational properties of reduced coda consonants [21][22]. In view of this, $V_L C_A V_R$ in English cannot be syllabified linearly and intervocalic consonants should be considered as a rather independent group in TTS synthesis.

References

[1] Muhammad Masud Rashid, Md. Akter Hussain, M. Shahidur Rahman, “Text normalization and Diphone preparation for Bangla speech synthesis”, *Journal of Multimedia*, vol. 5, pp. 551-557, 2010.

[2] Wang Lijuan., Zhao Yong and Chu Min, “Context-dependent boundary model for refining boundaries segmentation of TTS units,” *IEICE Transactions on Information and Systems*, v E89-D, n 3, pp. 1082-1091, 2006.

[3] Nicolas Dumay, Alain Content, “Searching for syllabic coding units in speech perception,” *Journal of Memory and Language*, vol. 66, pp. 680-694, May 2012.

[4] Mohammad Nurul Huda, Manoj Banik, Ghulam Muhammad, Mashud Kabir, Bernd J. Kröger, “Effects of Syllable Language Model on Distinctive Phonetic Features (DPFs) based Phoneme Recognition Performance”, *Journal of Multimedia*,vol. 5, pp. 543-550, 2010.

[5] Yan Deng, Wei-Qiang Zhang, Yan-Min Qian, Jia Liu, “Time-Frequency Cepstral Features and Combining Discriminative Training for Phonotactic Language Recognition”, *Journal of Multimedia,* vol. 6, PP. 178-183, 2011.

[6] Chunyi Guo, “Neighboring digits pattern training method in quickly-spoken connected mandarin digits speech recognition”, *Journal of Multimedia,* vol. 6, PP. 300-307, 2011.

[7] Huan Zhao, Xiujuan Peng, Lian Hu, Gangjin Wang, “An improved speech enhancement method based on teager energy operator and perceptual wavelet packet decomposition”, *Journal of Multimedia,* vol. 6, PP. 308-315, 2011.

[8] Content, A., Meunier, C., Kearns, R.K., Frauenfelder, U.H., “Sequence detection in pseudowords in French: where is the syllable effect?” *Language and Cognitive Processes*,,vol. 16, pp 609–636, 2001.

[9] Modarresi, G., Sussman, H., Lindblom, B., Burlingame, E., “Stop place coding: an acoustic study of CV, VC#, and C#V sequences”, *Phonetica*, vol. 61, pp. 2–21, 2004.

[10] Peggy P.K. Mok, “Effects of consonant cluster syllabification on vowel-to-vowel coarticulation in English”, *Speech Communications*, vol. 54, pp. 946-956, 2012.

[11] Selkirk, E., “The syllable, In Harry Van der Hulst and

Norval Smith (eds.) The structure of phonological representations", *Dordrecht: Foris*, Part II. pp. 337-383, 1982.

[12] Wells, J. C., "Syllabification and allomorphy", *In Studies in the pronunciation of English* (S. Ramsaran, editor), London: Routledge, pp. 76-86, 1990.

[13] Kahn, D., "Syllable-based generalizations in English phonology", *PhD dissertation*, MIT, New York: Garland, 1980.

[14] Hooper, Joan B., "Constraints on schwa-deletion in American English", *In J. Fisiak (ed.) Recent developments in historical phonology*, The Hague: Mouton. 183-207, 1978.

[15] D. Fallows, "Experimental evidence for English syllabification and syllable structure", *Journal of Linguistics*, vol. 17, pp. 309-319, 1981

[16] Pharao, N., "The Psychological Reality of Ambisyllabic Consonants in Danish", *Proceedings of the 15th ICPhs Barcelona*, Barcelona, 2003.

[17] Schiller, N. O., Meyer, A. S., & Levelt,W. J. M., "The syllabic structure of spoken words: Evidence from the syllabification of intervocalic consonants", *Language and Speech*, vol. 40, pp. 103–140. 1997.

[18] Treiman, R., Judith A. Bowey and Derrick Bourassa, "Segmentation of spoken words into syllables by English-speaking children as compared to adults", *Journal of Experimental Child Psychology*, vol. 83, pp. 213-238, 2002.

[19] Takasawa. M., "The Duration of the Consonants in Syllable-Final and Syllable-Initial Positions in Castilian Spanish", Proceedings of the 15th ICPhS Barcelona, Barcelona, 2003.

[20] R.J.J.H. van Son.; Jan P. H. can Santen. "Strong Interaction between Factors Influencing Consonant Duration", http://www.fon.hum.uva.nl/IFApublications/Eurospeech97/A0456/A0456.html

[21] Fougeron, C.; Bagou, O.; Content, A.; Stefanuto, M.; Frauenfelder,U. "Looking for acoustic cues of resyllabification in French",. *Proceedings of the 15th ICPhS Barcelona*, Barcelona, 2003

[22] Hazel Morton, Nancie Gunson and Mervyn Jack, "Attitudes to subtitle duration and the effect on user responses in speech interactive foreign language learning" , *Journal of Multimedia*, vol. 6, pp. 436-446.

© 2013 ACADEMY PUBLISHER