euronoise

Production and perception of vietnamese short vowels

Viet Son Nguyen[a], René Carré[b] and Eric Castelli[a]

[a]MICA center HUT - CNRS/UMI2954 INP Grenoble, C10 Hanoi university of Technology, No1 Dai Co Viet street - Hai Ba Trung, 0084 Hanoi, Viet Nam
[b]Laboratoire Dynamique du Langage, UMR 5596, CNRS, Université Lyon 2, 14 Avenue Marcelin Berthelot, 69363 Lyon cedex 07, France
viet-son.nguyen@mica.edu.vn

It is well known that vowels can be produced in isolation, acoustically stable, in such a way that they can be represented as points in the F1-F2-F3 space. Vietnamese language presents 13 vowels, but only 9 vowels [i, u, e, o, a, ɔ, ɛ, ɯ, ɤ] can be pronounced in isolated mode. A previous study showed that the 4 remaining vowels [ă, ĕ, ŏ, ɤ̆] have same target characteristics (F1, F2, F3) as, respectively, the vowels [a, ɛ, ɔ, ɤ] but their dynamics (the rates of CV transitions) are clearly distinct.

The paper analyses the production of Vietnamese VC, including classical vowels and special vowels in terms of duration, format evolution and rate of VC transitions. Measurements show that vowel durations of [ă, ŏ, ɤ̆] are always shorter than the one of the corresponding vowels [a, ɔ, ɤ] and are not acoustically stable, on the one hand, and the vowel [ĕ] has acoustic characteristics as a diphthong. In perception tests, synthesized syllable [a-t], [ɤ-t], [ɔ-k] with changing vowel duration are recognized as [ă-t], [ɤ̆-t], [ŏ-k] when then duration of initial vowel [a, ɤ, ɔ] are 50%-70% shorter. It means that the vowel duration is an important parameter that allows Vietnamese distinguishing the classical vowels and special vowels in Vietnamese language.

1 Introduction

The representation of vowels in the F1-F2 space was used by [1, 2, 3] and many others. It has been used extensively because of its simplicity and its explicative power. The first two formants (F1 and F2) are known to be necessary and in most instances sufficient for the representation of vowels. They denote the frequencies of the first two resonant modes of the vocal tract, and they permit intelligible synthesis of vowels.

In Vietnamese, linguists [4] affirm that there exists only 9 classical Vietnamese vowels 'a', 'e', 'ê', 'i', 'o', 'ô', 'ơ', 'u', 'ư' (written in Vietnamese characters), that correspond respectively to [a], [ɛ], [e], [i], [ɔ], [o], [ɤ], [u], [ɯ] (in phonetic), and four special Vietnamese vowels [ă], [ĕ], [ŏ], [ɤ̆] corresponding respectively to the vowels [a], [ɛ], [ɔ], [ɤ]. However, when it is asked to pronounce these vowels in isolation (with a flat monotonous pitch), Vietnamese answer that they could not.

In a previous paper [5], we showed that the two special vowels [ă] and [ɤ̆] present the same target characteristics (formants F1 and F2) as [a] and [ɤ]. However, they present different dynamic characteristics: non monotonous pitch and, in VC production, the transition rate is faster than for the other vowels. In perception, [5] found that the 11 Vietnamese vowels (9 classical vowels and 2 special vowels [ă] and [ɤ̆]) are clearly distinct.

Such four special vowels can be interesting to study because they are outside of hypotheses: what are their acoustic characteristics compared with characteristics of the other ones? How can they be perceived? In this paper, different measurement and perceptual experiments are proposed to study these hypotheses for Vietnamese vowels.

2 Structure of Vietnamese syllable.

According to studies of linguists, a Vietnamese syllable in its complete form has three parts: Initial part, final part and tone. The final part can be divided into three smaller components, i.e. medial part, nucleus part and ending part. So the full form of a syllable has five components: Initial part, medial part, nucleus part, ending part and tones (Fig.1)

<table>
<tr><td colspan="4">Tone</td></tr>
<tr><td rowspan="2">Initial part</td><td colspan="3">Final part</td></tr>
<tr><td>Medial</td><td>Nucleus</td><td>Ending</td></tr>
</table>

Fig.1 Structure of Vietnamese syllable [6]

The nucleus part is the centre of the syllable and is always a vowel. Base on articulation, the moving of the tongue and the aperture of mouth, Vietnamese vowels can be divided into: narrow vowels, wide vowels, and etc. In Vietnamese, there are 13 vowels (9 classical vowels and 4 special vowels) and 3 diphthongs [4]. The ending part is a sound standing at the end of the final part or the end of the syllable. In Vietnamese, there are six ending consonants, i.e. [p], [t], [k], [m], [n], [ŋ] and two ending semivowels [w], [j].

From point of view of vowel's duration [4, 7], there are: long vowels and short vowels. In the case of short vowels, Vietnamese cannot pronounce them in isolation as long vowels, and they cannot appear individually. The short vowels must always be combined with one of the ending part.

3 Vowel production

In order to study the Vietnamese vowels, a small Vietnamese vowel corpus was built by five Vietnamese subjects (5 males Son, Huy, Dat, Linh, Vinh who were born and live in the north of Vietnam). The 8 Vietnamese oral vowels [a], [ɛ], [ɔ], [ɤ], [ă], [ĕ], [ŏ], [ɤ̆] were pronounced 5 times in four sentences (5 times for each sentence): "Say V1C2 softly", "Say C1V1C2 softly", " Say V1V2 softly" and " Say C1V1V2 softly" where C1 is the initial consonant [b], V1 is one of the 8 Vietnamese vowels above; V2 is one of the two semivowels [w, j], and C2 is one of the 3 final consonants [p, t, k]).

The first two formants of the vowels were measured during all the productions of the quasi stable parts of the vowels. In Fig.2 the formant variations of each vowel are represented for two male subjects and one production. The

two sets of vowel [a, ă] and [ɤ, ɤ̆] are more or less acoustically closed in the plane F1-F2 [5]. It also points out that the set of vowel [ɔ, ɔ̆] acoustically closed. For the vowel [ĕ], the starting acoustic point is more or less closed to the vowel [a], but the ending point is closed to the vowel [ɛ] (Fig.3). From literature [8, 9], the special vowel [ĕ] is considered as more or less similar to a diphthong.

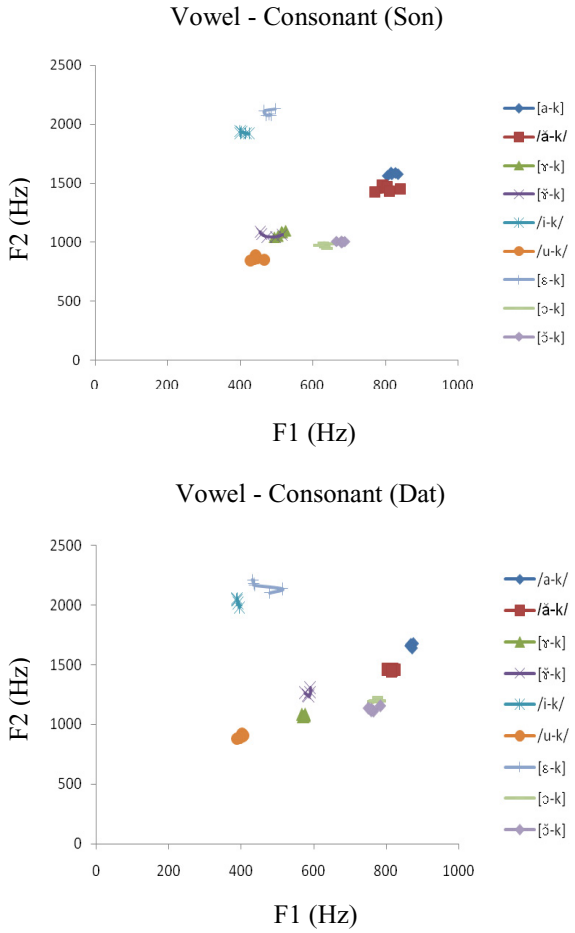

Fig.2 Representation in the F1-F2 plane of the vowel formant frequencies for two male subjects.

From point of view of vowel's duration, two classes of Vietnamese vowels are defined: long vowels (classical vowel) and short vowels (special vowel) [4, 7]. We intend to study if Vietnamese subjects can recognize and distinguish the classical vowels and special vowels by their duration.

3.1 Vowel characteristics in VC2 and C1VC2 context

The duration of vowel V was measured in all the production of their quasi stable states in VC2 and C1VC2 context. The durations of the vowels [a] and [ă] in both VC2 and C1VC2 context are represented for all of 5 males subjects in Table 1 and Table 2 (the mean value is calculated for 5 productions of each subject). In VC2 and C1VC2 context, the two special vowels [ĕ] and [ɔ̆] cannot be combined with consonant [p, t] (Vietnamese can not pronounce them).

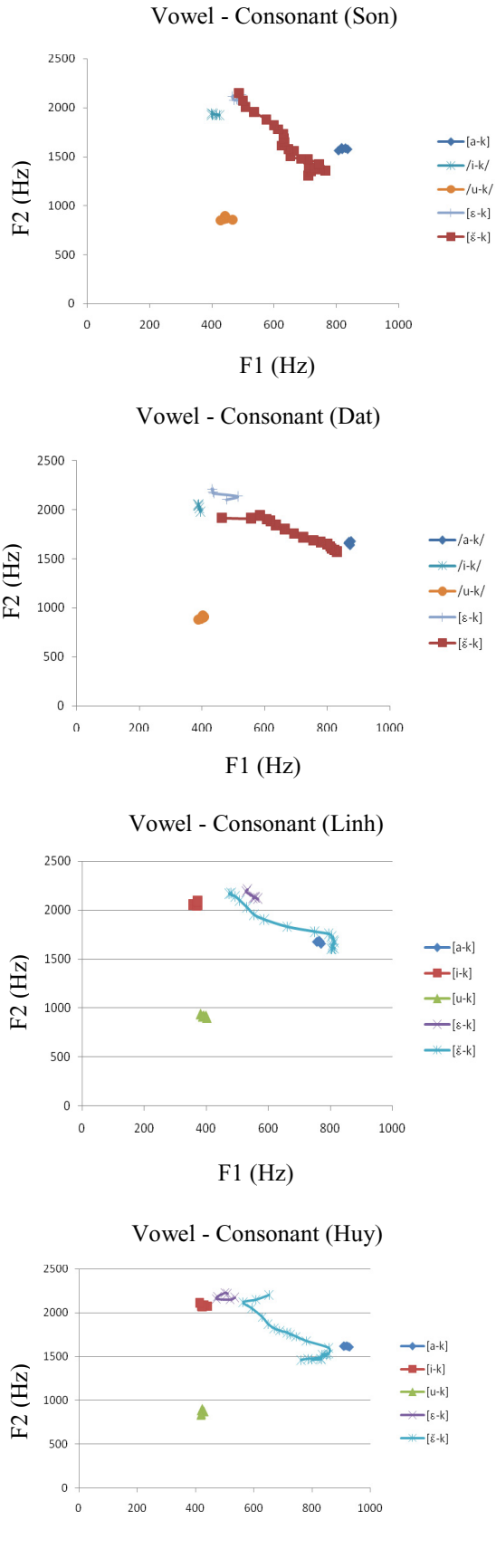

Fig.3 Representation in the F1-F2 plan of the consonant-vowel formant frequency for 4 male subjects and one production.

It is interesting to note that in all of the combinations in the both VC2 and C1VC2 context, the duration of vowel [a] is always longer than the one of the vowel [ă]. The mean

duration of vowel [a] is about 130ms in the VC2 context, and 115ms in the C1VC2 context, while the one of the vowel [ă] is 56% shorter (73ms) in the VC2 context and 53% shorter (62ms) in the C1VC2 context.
Table 3 represents the results for the other vowels (average for 5 subjects).

Subjects	[a] duration (ms)					
	VC2 context			C1VC2 context		
	[p]	[t]	[k]	[p]	[t]	[k]
NVSon	143	128	147.2	118.2	122.6	145.2
NQHuy	123	110.2	144.4	114.8	108.2	117.8
TDDat	113.2	120.2	143.6	100.4	98.4	129.4
TKLinh	132	144	165	121	110.2	124
NQVinh	80.8	126.8	141.2	100.8	105.6	115.6
Average	**126.56**	**125.84**	**148.28**	**111.04**	**109**	**126.4**

Table 1 Vowel duration in VC2 and C1VC2 context (V is classical vowel [a], C1 is [b] consonant, C2 is one of final consonants [p, t, k]).

Subjects	[ă] duration (ms)					
	VC2 context			C1VC2 context		
	[p]	[t]	[k]	[p]	[t]	[k]
NVSon	78.2	83.6	82	71.6	66.8	72.2
NQHuy	62.4	66.8	81.2	52.8	44.8	60
TDDat	54.6	65.2	70.2	51.2	47.8	60.6
TKLinh	80.8	75.6	75.2	78.8	61.6	69.6
NQVinh	80.8	68	82.4	64	65.6	59.2
Average	**71.36**	**71.84**	**78.2**	**63.68**	**57.32**	**64.32**

Table 2 Vowel duration in VC2 and C1VC2 context (V is special vowel [ă], C1 is [b] consonant, C2 is one of final consonants [p, t, k]).

Vowel	**Vowel duration (ms)**					
	VC2 context			**C1VC2 context**		
	[p]	[t]	[k]	[p]	[t]	[k]
[ɤ]	116.36	112.84	118.95	97.2	82.56	102.22
[ɤ̆]	68.8	66.84	77.2	56.16	55.68	60.04
[ɔ]	114.23	112.35	125.24	101.72	103.33	104.48
[ɔ̆]	None	None	79.44	None	None	65.2

Table 3 Mean vowel duration in VC2 and C1VC2 context (C1 is [b] consonant, C2 is one of consonant [p, t, k]).

In all vowels, and in the both VC2 and C1VC2 context, the duration of classical vowels ([ɤ], [ɔ]) is always longer than respectively the one of the special vowels ([ɤ̆], [ɔ̆]). The mean duration of vowel [ɤ̆] is 61% shorter than the one of vowel [ɤ] in both of the VC2 and C1VC2 context (71ms/116ms and 57ms/94ms, respectively). This result still remains the same in the case of the set of vowels [ɔ, ɔ̆] in both case of VC2 and V1CV2 context.
Furthermore, it can be note that in all cases of these sets of vowels as well as in both of the two contexts (VC2 and C1VC2), the effect of C2 ([p], [t], and [k]) on the vowel duration is not worth considering.

3.2 Vowel characteristics in context V1V2

We measured the duration of the vowel V1 in all the productions of the quasi stable parts in V1V2 and C1V1V2 context (C1 is [b] consonant and V2 can be one of two semivowels [w, j]). The durations of the vowels [a, ă] are represented for all of the 5 male subjects (the mean values are calculated for 5 productions of each subjects) in Table 4 and Table 5.

Subjects	[a] duration (ms)			
	V1V2 context		C1V1V2 context	
	[w]	[j]	[w]	[j]
NVSon	111.8	100.6	98.8	79.8
NQHuy	84.8	115.2	97.4	95.2
TDDat	120	91.2	105.6	73.6
TKLinh	133.8	124.8	119.8	106
NQVinh	101.2	95.2	118.4	80
Average	**110.3**	**105.4**	**108**	**86.92**

Table 4 Vowel duration in V1V2 and C1V1V2 context (V1 is classical vowel [a], C1 is [b] consonant, V2 is one of two semivowels [w, j])

Subject	[ă] duration (ms)			
	V1V2 context		C1V1V2 context	
	[w]	[j]	[w]	[j]
NVSon	49.4	52.8	62.4	31.2
NQHuy	57.6	57.6	51.2	36.8
TDDat	50.2	35.8	65.6	39.2
TKLinh	58.2	64	66.4	54.2
NQVinh	73.4	44.8	78.4	44.6
Average	**57.76**	**51**	**64.8**	**41.2**

Table 5 Vowel duration in V1V2 and C1V1V2 context (V1 is special vowel [ă], C1 is [b] consonant, V2 is one of two semivowels [w, j])

Once again, it is interesting to remark that the duration of the classical vowel [a] is longer than the one of the special vowel [ă] in all production.
In the V1V2 and C1V1V2 context, there are several cases in which V1 and V2 cannot be combined together (Vietnamese have not ability to pronounce them), for example: [ɤ-w], [ɛ-j], [ɛ̆-w], [ɛ̆-j], [ɔ̆-w], [ɔ̆-j], [ɔ-w].

Vowel	**Vowel duration (ms)**			
	V1V2 context		**C1V1V2 context**	
	[w]	[j]	[w]	[j]
[ɤ]	None	111	None	76.6
[ɤ̆]	68	54.2	58.04	44.24

Table 6 Mean vowel duration of vowel [ɤ] and [ɤ̆] in V1V2 and C1V1V2 context (C1 is [b] consonant, V2 is one of semivowel [w, j])

Making a comparison in the same context (V1V2 and C1V1V2 with the same V2), the duration of the vowel [ɤ] is always longer, 49% and 58% respectively, than the one of the vowel [ɤ̆] (Table 6).

4 Vowel perception

The perception of these two sets vowels (classical vowels and special vowels) is then studied. In the first part, the experiment consisted in searching the boundary of duration between the special vowel and classical vowel, respectively [a, ă], [ɤ, ɤ̆] and [ɔ, ɔ̆] by varying the vowel duration. The classical vowels [a], [ɤ] and [ɔ] are synthesized in the

context VC (C is [t] or [k] consonant) with the initial duration of 150ms. Ten subjects (5 males and 5 females) had to decrease the vowel duration to find out the boundary of the special vowel and the classical vowel for each of set of vowels. Results are represented in Table 7.

Subjects	Boundary (ms)			
	Sex	[at-ăt]	[ɤt-ɤ̆t]	[ɔk-ɔ̆k]
NVSon	M	75	53	87
TDDat	M	55	48	94
HVThai	M	126	78	86
NCPhuong	M	104	73	110
TKLinh	M	93	80	92
NLLan	F	103	61	96
BTThuy	F	110	50	106
LMThuy	F	83	79	81
LTLan	F	66	37	96
NTHuong	F	105	48	109
Average	M	90.6	66.4	93.8
Average	F	93.4	55	97.6
Average		92	60.7	95.7

Table 7 Results of search of the boundary of duration between the special vowel and classical vowel in VC context (C is [t] or [k] consonant)

This experiment indicated that there always exists a boundary in vowel duration between the special vowel and classical vowel for each set. All 10 subjects (males and females) always can find out the special vowels [ă], [ɤ̆], [ɔ̆] respective to the classical vowels [a], [ɤ] and [ɔ] by decreasing the duration of classical vowel. The result showed that the females recognize the special vowel later than the male, exceptionally in the case of the set of the vowels [ɤ- ɤ̆]. The boundary of set vowel [ɔ-ɔ̆] is the longest boundary (95.7ms) and the one of the set of the vowel [ɤ- ɤ̆] is the shortest. There is not noticeable effect of the consonantal context (with [t] consonant or [k] consonant).

In the second part, the perception of classical vowels and special vowels, respectively, in the V1V2 context was tested. All the available combinations of classical vowel with two semivowels [w, j] are synthesized ([a-j], [a-w], [ɤ-j]). Three durations were changed: V1 duration, V2 duration and V1-V2 transition duration. First, in the case of [a-j] syllable, we kept the V1 duration ([a] vowel) and V2 duration ([j] vowel) constant, 100ms and 150ms respectively, and we increased the V1-V2 transition duration from 100ms to 400ms by 10 steps. The test was carried out by 6 Vietnamese subjects (3 males and 3 females). Table 8 represents the results.

Trans dur (ms)	[a-j]	[a-V3-j]	[a-V3-V4-j]
100	100%	0%	0%
133	100%	0%	0%
166	50%	50%	0%
199	5.56%	94.44%	0%
233	0%	94.44%	5.56%
266	0%	72.22%	27.78%
299	0%	38.89%	61.11%
333	0%	22.22%	77.78%
366	0%	0%	100%
399	0%	5.56%	94.44%

Table 8 Results of the search of V1-V2 transition duration in V1V2 context.

Table 8 shows that Vietnamese can hear the third vowel as [ɛ] or [e] according to the transition duration. When the V1-V2 transition duration is enough long (from 166ms to 266ms), [ɛ] is recognized, and if we continue to lengthen the transition duration (longer than 299ms), Vietnamese can perceive [e]. These two vowels [ɛ] and [e] stay on the trajectory [a-i] in the F1-F2 plan, and from [a] to [i], we reach the vowel [ɛ] first. On the other hand, Vietnamese could not recognize the special vowel [ă] in this case. It is important to note that the transition duration in V1V2 context is not important to distinguish the special vowel [ă] and classical vowel [a].

Secondly, we kept the V1-V2 transition duration (50ms), V1 duration ([a] or [ɤ] vowel) and V2 duration ([w] or [j]) are remained equal to 50ms. Twenty Vietnamese listeners (10 males and 10 females, one time for each subject) listen to the synthesized syllables and have to select which syllable they hear [aj], [ăj], [aw], [ăw], [ɤj], [ɤ̆j] and NA (none acknowledged). The results are given in Table 9.

[a]-[j]			[a]-[w]			[ɤ]-[j]		
[aj]	[ăj]	NA	[aw]	[ăw]	NA	[ɤj]	[ɤ̆j]	NA
0%	100%	0%	31.58%	68.42%	0%	31.58%	63.16%	5.26%

Table 9 Results of search of vowel duration in V1V2 context

The results clearly show that the two special vowels [ă, ɤ̆] are recognized for the two cases of semivowel V2 ([w] and [j]) if their duration is less than a threshold. The subjects can realized the task for the set [a, ă] more easily than for the set [ɤ, ɤ̆]. In the case of [a, ă], it seems more difficult to distinguish if V2 is the semivowel [w].

5 Conclusions and Perspectives

Vietnamese vowel production shows three sets of vowels [a, ă], [ɤ, ɤ̆], [ɔ, ɔ̆] the targets of which are more or less acoustically closed in the plane F1-F2, but the set of vowels [ɛ, ɛ̆] is clearly difference. Vowel [ɛ̆] (with starting acoustic point more or less closed to the classical vowel [a] and the ending acoustic point closed to vowel [ɛ]) is more or less similar to a diphthong. However, it seems that, in all context (VC, C1VC2, and V1V2), the special vowels has duration shorter 50%-70% than the classical vowels. In perception, the boundary between these two sets of vowels was easily found. On the other side, the transition duration is not an important parameter to distinguish the special vowels and the classical vowels.

These results lead to emphasize the role of the duration of vowels as an important parameter that allows to Vietnamese to distinguishing the special vowels and the classical vowels in Vietnamese language. However, this parameter (vowel duration) can not still explain why Vietnamese could not pronounce the special vowel in isolation. The "dynamic hypothesis" in [5] maybe the first answer to this question.

In order to know better the characteristics of the sets of the Vietnamese special vowels, we plan to continue our experiments by studying the trajectories of the first three formants (F1, F2, F3) in both contexts (VC and V1V2) by modifying their dynamic characteristics, i.e. the transition duration between V and C, and the gradients of the first three formants.

References

[1] R. K. Potter, J. C. Steinberg, "Toward the specification of speech", Journal of the Acoustical Society of America 22, 807-820 (1950).

[2] G. E. Peterson, H. L. Barney, "Control method used in a study of the vowels", Journal of the Acoustical Society of America 24, 175-184 (1952).

[3] G. Fant, "Acoustic analysis and synthesis of speech with applications to Swedish", Ericsson Technics No. 1 (1959)

[4] T.T. Doan, "Ngữ âm tiếng Việt" (Vietnamese phonetics), Hanoi National University Publishing House (1977)

[5] E. Castelli, R. Carré, "Production and perception of Vietnamese vowels", Interspeech 2005, 2881-2884 (2005)

[6] H.Q. Nguyen, "Ngữ pháp tiếng Việt" (Vietnamese grammar), Encyclopedia Publishing House (2002)

[7] T. Hoang, M. Hoang, "Remarques sur la structure phonologique du vietnamien", Essais linguistiques-études vietnamiennes, $N^0$40-Ha Noi (1975)

[8] J.C. Catford, "A practical introduction to phonetics", Clarendon press Oxford, 115-116 (1988)

[9] L. Rabiner, "Fundamentals of speech recognition", Published by Prentice Hall, 28-29 (1993)