

## B-5

### 日本語アクセントが母音の長短の知覚に与えるトップダウン効果

有賀照道, 松原理佐

東京大学 総合文化研究科

ariga@phiz.c.u-tokyo.ac.jp, matsubara-risa761@g.ecc.u-tokyo.ac.jp

#### 要旨

音声単語認知はさまざまなレベルの情報が相互に関連することにより、音声を安定的に認知する頑健なシステムになっている。音声単語認知にはアクセントも重要な役割を果たし、ボトムアップに語の活性化を規定するだけでなく、音素のカテゴリー判断にトップダウン効果をもたらす。本研究では、日本語のピッチアクセントが母音の長短のカテゴリー判断に影響を与えるかを、同定課題によって検討した。「にんぎょ HLL—にんぎょう HLLL」（「にんぎょ HLL」（人魚）のみ有意味語）／「にんぎょ LHH—にんぎょう LHHH」（「にんぎょう LHHH」（人形）のみ有意味語）のようなペアを用いて、語末母音の母音の時間長を連続的に変化させたとき、母音の長短のカテゴリー判断はアクセント情報を考慮して語が有意味語となるようにバイアスが生じた。このことから、日本語のアクセントは母音の長短のカテゴリー判断にトップダウン効果をもたらすものであることが示唆された。

#### 1. 背景

人間の音声単語認知はさまざまなレベルの音声情報の処理が相互に関連することにより、発話された語を即時的に安定的に認知するシステムになっている。例えば TRACE モデル (McClelland & Elman, 1986) では、音響情報レベル、音韻（音素）レベル、単語レベルの情報が並列的に処理され、各レベルの処理が他のレベルの処理に相互に影響をもたらす。音韻レベルの情報処理は高次の単語レベルの処理にボトムアップ効果をもたらすことで、音韻情報に適合する語の認知を促進する。一方で、単語レベルの情報処理は低次の音韻レベルの処理にトップダウン効果をもたらし、音韻レベルの知覚を逆に促進する役割を持っている。

このトップダウン効果の存在により、音声単語認知メカニズムは情報の欠損に対して頑健なメカニズムになっている。例えば、単語中の一部の音素を無音や雑音に置き換えて欠落させたとき、失われた音素情報は前後の文脈をもとに補われて知覚される (Warren, 1970)。さらに、語提示の環境下において機械的に音響情報を加工することにより曖昧になった音素が入力されたとき、その音素のカテゴリー判断 (Liberman et al., 1957) は、語全体が有意味語となるようにバイアスがかかる (Ganong, 1980; McQueen, 1991)。これらのメカニズムは、常に完全な状態で入力が行われるとは限らない音声言語理解の状況下で、安定的に語の認知を達成するための知覚的恒常性の一環と捉えることができる。

音声単語認知には、分節音の処理と同時に、プロソディの情報も重要な役割を果たすことが知られる (Cutler et al., 1997)。日本語ではピッチアクセント（以下、単にアクセントとする）が語彙的に指定され、語彙情報の一部をなす。誤ったアクセントで語が提示されると、その語が正しく認知される割合は低下する (Minematsu & Hirose, 1995; Slowiczek, 1990)。また、「じどう HLL（児童）」／「じどう LHH（自動）」のような同音異アクセント語は、そのアクセント情報をもとに弁別的に語アクセスが行われる (Ariga, 2022; Cutler, 1986; Sekiguchi & Nakajima, 1999)。これらの研究は音声単語認知においてアクセント情報は目的の語の活性化にボトムアップに影響を与えることを示すものである。

アクセント情報はさらに、低次の音素レベルの情報処理にもトップダウン効果をもたらす。Connine et al. (1987) は *digress-tigress* のような英語のストレスペアにおいて、語頭子音 (/d/または/t/) の VOT を連続的に変化させたときの有声性のカテゴリー判断を検討した。このペアは語のストレスと語頭子音の

有声性が連動しており、*digréss* は有意味語であるが\**tigréss* は無意味語である。同様に、*tigréss* は有意味語であるが\**digréss* は無意味語である。結果、語頭子音のカテゴリ判断は *digréss* から\**tigréss* へ連続的に VOT を変化させたときには/d/と判断する方向にバイアスがかかり、\**digréss* から *tigréss* に連続的に VOT を変化させたときには/l/と判断する方向にバイアスがかかった。これは語頭子音の有声性の判断という音素レベルの情報処理が語全体が有意味語となるように影響を受ける (Ganong, 1980) 際に、語全体の有意味性がアクセント情報を考慮して計算されることを示唆しており、アクセント情報が音素の知覚にトップダウン効果をもたらすと解釈することができる。

同様の効果は日本語のアクセントペア (e.g., 「たんす LHH」 / 「ダンス HLL」) においても観察される (Ariga & Matsubara, to appear)。日本語のアクセントは方言間の多様性 (窪田 2021) や複合語アクセント、平板化などの音韻規則 (Kawahara, 2015) により、音声単語認知において常に目的の語の認知を促進する要因とはなりがたい (柴田 1961)。それにもかかわらず、低次の音素レベルの情報処理において、曖昧な音素の入力を補う情報として日本語アクセントが重要な役割を果たすことが示唆される。

しかし、このアクセントがもたらすトップダウン効果が、音素のカテゴリ知覚以外の情報処理に影響をもたらすかどうかは明らかになっていない。先行研究 (Ariga & Matsubara, to appear; Connine et al., 1987) において検討されたのは VOT の時間長を操作することによって得られる有声性の対立のカテゴリ判断にとどまる。

日本語には、母音の長短の対立がある (e.g., 「おばさん」 / 「おばあさん」)。母音の長短は母音の物理的時間長を主な手がかりとして判断され、日本語母語話者は短母音と長母音を容易に弁別することができる (Dupoux et al., 1999)。日本語の語彙には、母音の長短とアクセント型が語の有意味性に交絡する、すなわち、有意味語の範囲内で語を検索した際に、アクセント型を指定すれば母音の長短が連動して決まる語のペアが存在する (e.g., 「にんぎょ HLL」(人魚) / 「にんぎょう LHHH」(人形))。

このとき、母音の長短の知覚がアクセント情報によるトップダウン効果を受けるならば、母音の物理的時間長を連続的に変化させた場合に、その長短のカテゴリ判断はアクセント情報を考慮して語全体が有意味語となる方向にバイアスを受けると予想される。一方、アクセント情報は母音の長短の知覚にはトップダウン効果を与えないならば、母音の長短のカテゴリ判断は母音の物理的時間長のみによって決定され、アクセント型によるバイアスは生じないと考えられる。本研究ではこの可能性を、先行研究 (Ariga & Matsubara, to appear; Connine et al., 1987) にならった同定課題によって検討した。

## 2. 実験

### 2.1 実験刺激

実験には表 1 に示した、6 組の単語のペアを用いた。これらは 3 または 4 モーラの有意味語で語末の母音に長短の対立があり、かつアクセント型に頭高型 / 平板型の対立があるペアである。日本語語彙データベース (天野・近藤 1999) を参照して、これらの単語は東京方言においてその他のアクセント型で発音されることがなく、高親密度の語であることを確認した。

これらのペアにおいて、語末の母音の長短とアクセント型は語の有意味性と連動している。例えば、「にんぎょ HLL」(人魚) / 「にんぎょう LHHH」(人形) は有意味語であるが、そのアクセント型のみを入れ替えた「にんぎょ LHH」 / 「にんぎょう HLLL」という語は無意味語である。すなわち、語の有意味性の観点からは、アクセント型が指定されれば語末の母音の長短もそれに応じて決まる。したがって、アクセント情報が語末の母音の長短の知覚にトップダウンに影響を与えるとすれば、語末母音の長

表 1 実験刺激

短母音バイアス条件	長母音バイアス条件
あくしゅ HLL (握手)	あくしゅう LHHH (悪習)
げきど HLL (激怒)	げきどう LHHH (激動)
にんぎょ HLL (人魚)	にんぎょう LHHH (人形)
さっか LHH (作家)	サッカーHLLL
はなが LHH (版画)	ハンガーHLLL
ほっけ LHH (魚の名前)	ホッケーHLLL

表 2 語末母音の物理的時間長 (ms)

短母音バイアス 条件	時間長の比						長母音バイアス 条件	時間長の比					
	1.0	1.2	1.4	1.6	1.8	2.0		1.0	1.2	1.4	1.6	1.8	2.0
あくしゅ HLL	94	111	125	144	162	177	あくしゅう LHHH	100	116	132	149	170	188
げきど HLL	88	106	126	136	153	165	げきどう LHHH	105	126	148	176	190	215
にんぎょ HLL	128	134	158	171	194	204	にんぎょう LHHH	146	160	178	197	217	231
さっか LHH	105	128	151	168	185	193	サッカーHLLL	97	114	135	155	170	187
はなが LHH	126	145	168	187	210	232	ハンガーHLLL	111	140	153	171	206	230
ほっけ LHH	118	136	158	186	209	231	ホッケーHLLL	103	116	130	147	164	181

短のカテゴリ判断はアクセント情報を考慮して語が有意味語となる方向にバイアスが生じると予想される。

実験に用いたペアのうち、アクセント情報を固定した際、語末の母音が短母音である場合に有意味語となる語を短母音バイアス条件（にんぎょ HLL—にんぎょう HLLL）、語末の母音が長母音である場合に有意味語となる語を長母音バイアス条件（にんぎょ LHH—にんぎょう LHHH）とする。アクセント型が潜在的に与える効果を排除するため、ペア間で片方の条件の語に割り当てられるアクセント型はカウンターバランスをとった。

音声刺激は次のような手続きで作成した。まず、日本語東京方言母語話者（女性）による、語末が長母音である語の 2 種類のアクセント型での発音（e.g., 「にんぎょう HLLL」と「にんぎょう LHHH」）を収録した。音声は防音室において、量子化ビット 16 bit、標本化周波数 44.1 kHz で Praat 6.2.14 (Boersma & Weenink, 2022) を用いて録音した。この 2 種類の自然な発話から、語末の長母音の区間（/ningyoo/の /oo/の部分）を切り出し、Praat の Lengthen (overlap-add) 機能を用いて、母音の長さが長いものから順に比が 2.0, 1.8, 1.6, 1.4, 1.2, 1.0 の 6 段階となるように時間長を編集した。この方法で長さを調整した語末母音区間の音声をもとの音声ファイルの語末母音以前の部分とつなげることにより、語末母音が長母音から短母音へと連続的に変化する 6 段階の音声ファイルを作成した。ただし、語末母音と語末母音以前の部分のファイルを結合する際に、不自然さが生じないよう語末母音のはじめの 2-3 周期をカットしたため、時間長の比は厳密に等間隔にはなっていない。各刺激の語末母音の区間の時間長を表 2 に示した。母音母音の時間長が最も短い刺激と最も長い刺激について、それぞれ短母音、長母音として十分な聴覚印象をもつことを実験者（日本語母語話者）が確認した。

## 2.2 参加者

日本語東京方言母語話者 19 名（平均年齢 24.49 歳 ( $SD = 6.07$ )) が実験に参加した。実験参加者には所要時間相当額の謝金を支払った。

## 2.3 手順

実験はオンライン実験プラットフォーム PCibex (Zehr & Schwarz, 2018) を使用して、オンラインで行われた。参加者は各自のパソコンを用いて、自宅等の任意の静かな環境で課題を実施した。音声刺激はヘッドホンやイヤホンを用いて、適切な音量で聴取するよう教示した。

実験は 2 肢強制選択 (2AFC) 法に基づく、語末母音の長短の同定課題である。各試行の開始時に、画面の中央に注視点 (+) を 1000 ms の間提示した。その後、語末母音の長さが 6 段階に調整された単語音声の刺激が提示された。音声の流れ終わった直後、注視点が消え、画面上には流れた単語音声の候補をひらがなで表した 2 つの選択肢 (「にんぎょ」と「にんぎょう」) が提示された。参加者は、音声の語末母音の長短を判断し、その音声提示された選択肢のどちらに該当するかを判断し、画面上のボタンまたはキー押しにより回答した。この際、本実験の課題は参加者の言語能力を測るテストではなく、必ずしも正解があるわけではないことを教示において説明し、直感的に聞こえたと思う方を選択するように指示した。画面上の選択肢は参加者ごとに左右を入れ替えてカウンターバランスをとった。回答に時間制限は設けなかったが、2 つの選択肢のどちらかを選ばなければ次の試行に進めないようにした。参加者の回答後、1000 ms の間を空けて次の試行を開始した。

実験刺激は 2 つのバイアス条件×6 段階の母音の時間長×6 アイテムの合計 72 試行からなる。課題の教示を行い、練習試行 12 試行 (「かいご HLL」(介護) / 「かいごう LHHH」(会合) ペアを使用) を行った後、本試行を行った。本試行では 72 試行をランダムな順序で提示し、途中で 1 回任意の休憩をさせた。

同定課題が終わった後、課題で使用したアイテムに関する発音課題を行った。実験で使用した語 (表 1) を 1 つずつ画面に表示し、参加者に自身の知識に基づいて自然なアクセントで発音をさせた。この課題は実験参加者が実験アイテムに関して東京方言の適切なアクセント型の知識を持っているかを確認する目的で行われた。実験所要時間は合計で 10–20 分程度であった。

## 2.4 結果

発音課題の回答を実験者が確認し、参加者が東京方言の適切なアクセント型で発音していないアイテムがあった場合に、その参加者の当該アイテムを使用した試行のデータポイントを除外した。この操作により全体の 2.2 % のデータを除外した。参加者の回答は、語末の母音が長母音であると回答した場合 1、語末の母音が短母音であると回答した場合は 0 とコード化した。データ分析には R 4.2.1 (R Core Team, 2022) を使用した。

各バイアス条件、母音の時間長ごとに、参加者が語末母音を長母音であると回答した割合を図 1 に示す。短母音バイアス条件、長母音バイアス条件それぞれについて、母音の時間長ごとの回答率の変化をロジスティック回帰曲線によって示した。母音の長短の判断が切り替わる母音の時間長の比の 50 % 閾値は、短母音バイアス条件が 1.76、長母音バイアス条件が 1.37 (いずれも長母音として発話したもとの刺激を 2.0 としたときの長さの割合を示す) であった。

参加者の回答を一般化線形混合効果モデル (GLMM) を用いて分析した。参加者の回答 (0 または 1)

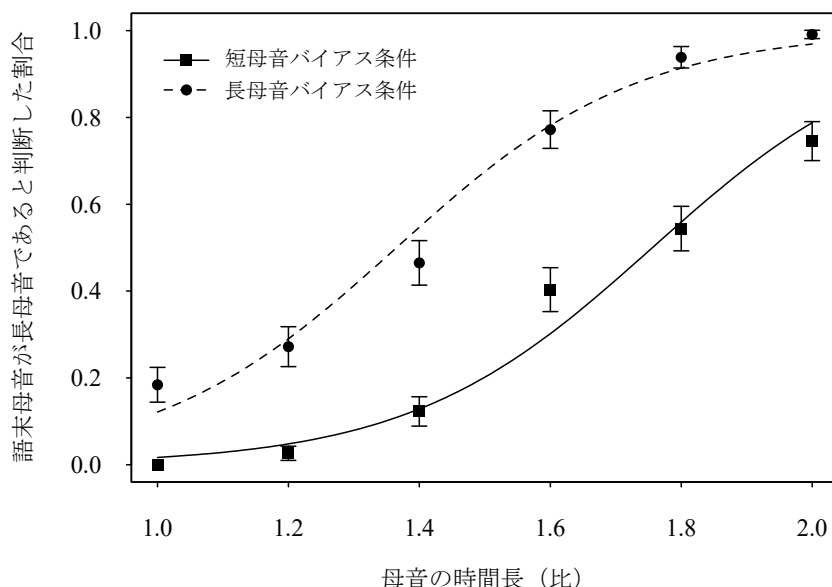


図 1 各バイアス条件，母音の時間長（比）ごとに，実験参加者が語末母音が長母音であると判断した割合。エラーバーは標準誤差（SE）を示す。実線は短母音バイアス条件（e.g., にんぎょ HLL—にんぎょう HLLL）のロジスティック回帰曲線，破線は長母音バイアス条件（e.g., にんぎょ LHH—にんぎょう LHHH）のロジスティック回帰曲線をそれぞれ示す。

表 3 GLMM による固定効果の推定結果

	$\beta$	SE	z	p	
(Intercept)	-8.887	0.510	-17.437	< .001	***
バイアス条件	2.127	0.936	2.273	.023	*
母音の長さ	11.398	0.621	18.363	< .001	***
バイアス条件×母音の長さ	0.123	1.187	0.104	.918	

Significant codes: \*\*\*  $p < .001$ , \*  $p < .05$

を従属変数とし，各バイアス条件（短母音バイアス条件を-0.5，長母音バイアス条件を 0.5），語末母音の時間長の比（1.0，1.2，1.4，1.6，1.8，2.0：値が大きいほど長母音としてカテゴリー化されると予測される），その交互作用を固定効果に，実験参加者とアイテムの個体差をランダム効果に設定した。

一般化線形混合効果モデルによる各条件の固定効果の推定値を表 3 に示す。バイアス条件 ( $p = .023$ ) と語末母音の時間長 ( $p < .001$ ) に有意差が見られたが，その交互作用 ( $p = .918$ ) は有意ではなかった。音声刺激の語末母音の物理的時間長が長くなるにつれ参加者が母音を長母音とカテゴリー化する割合は上昇したが，短母音バイアス条件に比べて長母音バイアス条件の方がより短い物理的時間長の刺激に対して長母音と回答する割合が高かった。すなわち，提示された音声のアクセント情報が，母音の長短のカテゴリー判断にバイアスをもたらしたと考えられる。

### 3. 考察

本研究では，音声単語認知において日本語のアクセントが母音の長短のカテゴリー判断にトップダウン効果をもたらすかどうかを検討した。「にんぎょ HLL」（人魚）／「にんぎょう LHHH」（人形）のペ

アにおいて、音声のアクセントが頭高型である場合、語末母音の時間長を連続的に変化させた際に語末母音は長母音よりも短母音とカテゴリー化される方向にバイアスが生じた。一方で、音声のアクセントが平板型である場合には、語末母音は短母音よりも長母音とカテゴリー化される方向にバイアスが生じた。これは日本語の頭高型アクセントの語彙の中に「にんぎょ HLL」は存在するが「\*にんぎょう HLLL」は存在せず、また平板型アクセント語彙の中に「にんぎょう LHHH」は存在するが「\*にんぎょ LHH」は存在しないことにより、聴取した音声が有意味語となるように語末母音の長短のカテゴリー判断を行ったことによるものと解釈することができる。

この結果は、音声刺激の語末母音の長短の判断には、単語のアクセント情報によるトップダウン効果が生じることを示唆する。すなわち、長短が曖昧な母音の入力を受けた際に、日本語話者はアクセント情報を参照して目的の語を推定し、聴取した語が有意味語となるように母音の長短を逆算する。音声単語認知において、アクセント情報は単なる音素のカテゴリー化 (Ariga & Matsubara, to appear; Connine et al., 1987) に影響を与えるだけでなく、母音の長短のカテゴリー化にも影響を与えられられる。

このことから、音声単語認知において母音の長短の処理とアクセントの処理には階層関係が存在することが示唆される。「\*にんぎょ LHH」や「\*にんぎょう HLLL」という誤った音声提示されたとき、音素レベルの情報として得られる母音の長短の情報がアクセント情報よりも優先的に語の解釈に用いられるとすると、これらの語はアクセントの誤りとして「にんぎょ HLL」「にんぎょう LHHH」として解釈されるはずである。しかし、本実験の結果にもとづけば、母音の長短の情報よりもアクセント情報が優先的に語の解釈に用いられ、アクセントではなく母音の長短が誤っていると見なされて「にんぎょう LHH」「にんぎょ HLL」のように母音の長短の情報の方が修正されて語が解釈されることになる。母音の長短という音素レベルに与えられる情報よりも、アクセントという単語全体にかかるプロソディ情報の方が、音声単語認知において信頼性の高い情報として優先度が高く処理されられると考えられる。

先行研究 (Ariga & Matsubara, to appear; Connine et al., 1987) では、アクセント情報がトップダウン効果をもたらす対象は個々の音素の種類判断であった。しかし、本研究の結果はアクセント情報が個々の音素の種類ではなく、その音素の質 (短母音か長母音か) の判断に影響を与えているという点に新規性がある。従来の音声単語認知モデル (TRACE モデル: McClelland & Elman, 1986) では、単語レベルの情報処理がトップダウン効果を与える先である音韻レベルの処理は音素の種類判断を行い、母音の長短といった音素の質の判断までは想定されていなかった。母音の長短の判断は、その母音の物理的時間長によって決定するが、語全体の話速といった超分節情報とも関連する (Hirata, 2004) 点で単なる分節音の処理とは異なると考えられる。そのため、本実験の結果を音声単語認知モデルの枠組みで説明するにはさらなる検討が必要である。

しかしながら、本実験の結果は、日本語のアクセントが語レベルの情報として低次のレベルの情報処理にトップダウンに影響を与えるものであることを示唆する。個々の音素情報に比べれば、単語レベルに与えられるアクセントの情報は単語認知の中で時間的に相対的に失われにくい。音素とともに並存するアクセントが音声情報に冗長性をもたらし、情報を補完することによって、曖昧な音素情報の入力を解釈することが可能になる。したがって、日本語の音声単語認知においてアクセント情報は語を安定的に認知するためのヒントとして重要な役割を果たしていると考えられる。

## 参考文献

天野成昭・近藤公久 (1999) 『NTT データベースシリーズ日本語の語彙特性』東京：三省堂。

- Ariga, T. (2022). Pitch accent constrains lexical activation in Japanese spoken word recognition: A semantic priming study. *Language and Information Sciences*, 20, 1–17.
- Ariga, T., & Matsubara, R. (to appear). Top-down effects of lexical pitch accent on phonetic categorization in Japanese. *The 20th International Congress of Phonetic Sciences*.
- Boersma, P., & Weenink, D. (2022). Praat: Doing phonetics by computer. <https://www.praat.org/>
- Connine, C. M., Clifton, C. E., & Cutler, A. (1987). Effects of lexical stress on phonetic categorization. *Phonetica*, 44(3), 133–146.
- Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, 29(3), 201–220.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech*, 40(2), 141–201.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion?. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1568–1578.
- Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110–125.
- Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, 32(4), 565–589.
- Kawahara, S. (2015). The phonology of Japanese accent. In: H. Kubozono (ed), *The Handbook of Japanese Phonetics and Phonology*. De Gruyter Mouton, 445–492.
- 窪蘭晴夫 (2021) 『一般言語学から見た日本語のプロソディー』 東京：くろしお出版。
- Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, N. C. (1957). The discrimination of speech sounds within and across phonemic boundaries. *Journal of Experimental Psychology*, 54(5), 358–368.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86.
- McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 17(2), 433–443.
- Minematsu, N., & Hirose, K. (1995). Role of prosodic features in the human process of perceiving spoken words and sentences in Japanese. *Journal of the Acoustical Society of Japan (E)*, 16(5), 311–320.
- R Core Team (2022). R: A language and environment for statistical computing. <https://www.r-project.org/>
- Sekiguchi, T., & Nakajima, Y. (1999). The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research*, 28(4), 439–454.
- 柴田武 (1961) 「日本語のアクセント」『言語生活』117, 14–20.
- Slowiczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech*, 33(1), 47–68.
- Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, 167, 392–393.
- Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). <https://www.pcibex.net/>