CLICK DETECTION IN ITALIAN AND ENGLISH

Brit van Ooyen[1,2], Anne Cutler[1] and Pier Marco Bertinetto[3]

[1]*MRC Applied Psychology Unit, 15 Chaucer Rd., Cambridge, CB2 2EF, United Kingdor*
[2]*Phonetics Laboratory, University of Leiden, P.O. Box 9515, Leiden, The Netherlands.*
[3]*Scuola Normale Superiore. 7 Piazza dei Cavalieri, 56100 Pisa. Italy*

ABSTRACT

We report four experiments in which English and Italian monolinguals detected clicks in continous speech in their native language. Two of the experiments used an off-line location task, and two used an on-line reaction time task. Despite there being large differences between English and Italian with respect to rhythmic characteristics, very similar response patterns were found for the two language groups. It is concluded that the process of click detection operates independently from language-specific differences in perceptual processing at the sublexical level.

Keywords: Speech perception, English, Italian, vowel duration, reaction time.

1. INTRODUCTION

The click detection paradigm has been popular with psycholinguists for investigating perceptual segmentation of sentences [1,2,3,4,5]. Two findings appear to be robust. First, clicks are more often than not displaced from their objective locations in the acoustic signal to the boundaries of perceptual units, notably in between clauses. Thus, click migration has been interpreted as being a species of the well-known phenomenon of perceptual closure, the tendency of perceptual units to resist interruption. Second, clicks are persistently perceived as having occurred before their objective location. Ladefoged and Broadbent [6] explain this tendency by reference to Titchener's law of prior entry [7]. This law states that, if a subject expects a certain stimulus, such as a click, along with some other unknown stimuli, such as an unpredictable string of words, the expected stimulus will be perceived faster than the unexpected stimuli. Ladefoged and Broadbent also found that errors in click location were larger and more frequent in sentential than in non-sentential word sequences, suggesting that toe size of perceptual units is larger in the processing of sentences than in the processing of non-sentential stimuli. Thus, click detection has yielded insight into the perception of comparatively large linguistic units.

However, linguistic units also exist at much lower levels, such as that of the syllable. Syllables can either be stressed or unstressed. This alternation in the speech stream is an important determinant of perceived rhythm; there is evidence from several languages that rhythm guides listeners' segmentation of the ongoing speech signal [8,9,10]. However, within languages that share the property of lexical stress, there can be considerable difference in the way in which the realisation of stress affects the perception of linguistic rhythm. English and Italian are a good example of such a difference: both have lexical stress, but different rhythms. In English, stressed syllables (and hence stressed vowels) tend to be much longer than unstressed ones. Unstressed syllables in continuous speech can undergo substantial shortening, which may even include a change in vowel quality, usually from a full vowel to schwa. In Italian, on the other hand, the duration of single segments is not quite as flexible. Though, like English, stressed syllables are usually longer than unstressed ones, the difference is smaller. By and large, Italian vowels preserve their intrinsic duration, irrespective of accentual status. Accordingly, Vekas and Bertinetto [11] propose to distinguish these two language types. They describe Italian as an example of a controlling language, and English as an example of a compensating language.

There is evidence that English listeners are relatively insensitive to durational compression of unstressed vowels. In a study comparing the performance of both English and Italian natives, Bertinetto and Fowler [12] found that Italian subjects are more sensitive to artificially shortened unstressed vowels than are English subjects. That is, the manipulated short vowels seemed more often unnatural to Italian listeners.

Given this language-specific sensitivity to perceived duration, a possible hypothesis is that the process of click detection also yields language-specific response patterns. Since syllables form perceptual units, and stressed syllables are comparatively long in English, it could be that clicks in stressed vowels are more accurately perceived than clicks in unstressed ones. In contrast, the comparative immunity from stress-induced durational flexibility as exhibited by Italian syllables may result in a smaller difference in click location performance for Italian listeners on stressed vs. unstressed syllables.

If vowel duration affects detection performance, then Italian

subjects should show the same degree of accuracy regardless of accentual status of a syllable. But if, on the other hand, click perception operates independently of syllable duration, we will find similar effects of stress in the two languages. Similarly, if click perception is sensitive to differences in perceived rhythm caused by differences in syllable durations, we should observe different response patterns for the two languages in a reaction time task. If not, performance will again be similar. Indeed, this last pattern was recently observed by Cutler *et al.* [13], in a study comparing the performance of English and French listeners. Both in a location task and in a reaction time task, these two languages yielded very similar results. The authors concluded that click detection is performed at a relatively low level and does not interact with sentence processing. However, before discarding click detection as a research tool altogether, a final cross-linguistic comparison is warranted. In this, two factors need to be controlled for whilst varying syllable duration:lexical stress and phonemic environment of the target items. Once these criteria have been met, we will be able to claim that any differences we observe in click detection performance can only be due to differential linguistic rhythms. But if we find similar response patterns in two languages with such different rhythmical structure, we will have to conclude that click detection does not tell us much below the level of sentence processing. As will be seen from the materials, English and Italian form a perfect pair for this purpose.

2. METHOD

Materials In each language, twenty four words were chosen in cognate pairs. Eight such pairs had primary stress on the first syllable *(lyrical-lirico)*, 8 on the second *{analysis-analisi)* and 8 on the third syllable *(metamorphosis-metamorfosi)*. The targets occurred in carrier sentences which were balanced such that, for each English-Italian pair, both the syntactic complexity and the rhythmic structure of the sentence up to the target word was the same. Table 1 lists all 24 experimental sentences. A further 10 sentences per language served as fillers.

Subjects Forty eight monolingual Italian students and 48 monolingual English students participated in the experiments. Half of the subjects performed the location task, the other half the reaction time task.

Experimental design The materials were recorded by a native male speaker of Italian and a native male speaker of British English, at a fluent speech rate and with normal sentence intonation. Each sentence was digitised and duplicated. A click was then placed in each copy, once in the syllable with primary stress and once in the preceding unstressed syllable in *analysis* and *metamorphosis* type words, and in the following unstressed syllable in *lyrical* type words. The click consisted of a 6 mstrain of 5 kHz square wave pulses and was located exactly in the middle of the vowel. In the filler sentences, clicks were located at word boundaries, in monosyllabic words and in polysyllabic

Table I. The 24 experimental sentences for English and Italian

The budget had been in a *critical* condition	Antonio gli fece una *critica* severa
The generous inheritance brought about a *metamorphosis* in the youngster	La grossa eredita portera con se una *metamorfosi* nella giovane
A serious accident caused a *paralysis* of both his legs	Un grave incidente causo la *paralisi* di entrambe le gambe
The physician said he was working on a thesis on the *retina* of the monkey	Il dottore stava scrivendo la sua tesi sulla *retina* della scimmia
His philosophy was so *theoretical* and difficult to understand	Il suo approccio era cosi *teoretico* e difficile da capire
The shortcomings of the computer were *electrical,* so we were told by the technician	Il guasto del nostro computer era *elettrico,* comme ci disse il tecnico
All his poetry was very *lyrical* and romantic	Il poem a era abbastanza *lirico* e seducente
Agents carried out an *analysis* of the situation	Pietro svolse poi un'*analisi* della situazione
The latest car is more *economical* than the old one	La nuova auto e piu *economica* della vecchia
The early *formula* was considered erroneous	La prima *formula* fu stimata scorretta
The modern rites are highly *esoteric,* there's no doubt about that	I nuovi riti erano *esoterici,* non c'e alcun dubbio
The proposal for the expansion was *symmetrical* and very simple	Il progetto dell'espansione era *simmetrico* ed alquanto semplice
The swing of *the pendulum* was irregular	I moti del *pendolo* sono isocroni
Fulbright was a well-known *philanthropist,* who lived in America	Fullbright fu un'noto *filantropo,* che visse negli USA
Although the speaker is *monolingual,* he understands a lot	Benche il soggetto sia *monolingue,* comprende quasi tutto
The location had so many *typical* characteristics	La citta aveva molte *tipiche* caratteristiche
The answer was very *political,* but it wasn't a he	La mossa fu molto *politica,* nondimeno fali
That event was the most *catastrophic* mat I can remember	Quell'evento fu il piu *catastrofico* che si ricordi
The company's working on the *synthesis* of the human voice	La ditta lavora sulla *sintesi* della voce umana
Gerry was famous for his *fantastic* imagination	Giorgio era noto per la *fantastica* cortesia
The curve is *parabolic* and easy to calculate	La curva e *parabolica* e di facile calcolo
The reporter had become *cynical* after his work had been repeatedly rejected	Lo scrittore divento *cinico,* quando il suo lavoro fu ripetutamente rifiutato
The treatment was merely *symptomatic,* and did not produce a lasting cure	La cura era solo *sintomatica,* e non produsse risultati durevoli
Judy gave him the *canonical* response that everyone expected	Giorgio diede la *canonica* risposta che tutti aspettavano

words. For each language one experimental tape was made. Each tape started with 5 filler sentences, followed by two renditions of each of the 24 experimental sentences, one for each click location, plus two renditions of the remaining 5 filler sentences. For each sentence, one rendition occurred in the first half and one in the second half of the experiment. Each sentence was preceded by a warning tone consisting of a 200 ms train of a 500 Hz sine wave.

Procedure In both experiments, subjects listened to the sentences in monaural presentation. For the location task, written instructions told the listeners that they would hear warning tones followed by sentences in their right ear, and that somewhere in each sentence a click would occur, which would be audible in their left ear. Their task was to judge the occurrence of this click as accurately as possible by marking its location with a vertical line on their response sheet. The response sheets contained the sentences, printed out with a blank space occurring between each phonological segment For example, in the sentences where *pendulum-pendolo* was the cognate pair, the texts looked like this:

The swing of the pendulum was irregular

Im otidel pend olosonois ocroni

This layout served to discourage listeners from concentrating on the boundaries between words, which are normally the only points associated with blanks in written texts. No subjects reported difficulty in using the response sheets. The subjects could indicate perceived click locations by drawing a vertical line either exactly through a phonological segment, or through a blank. They were also instructed not to look at a sentence before they had heard it, by keeping the page covered with a piece of card and moving this down to uncover the sentences one at a time. After each sentence the tape was stopped while the subject responded. In order to ensure that the subjects were attending to the content of the sentences, a short comprehension test was given afterwards.

For the reaction time task, the listeners were requested to press a response button immediately upon perceiving the click in their left ear. In order to ensure that attention was paid to the content of the sentences, listeners were asked to repeat the word on which they thought the click had occurred immediately after each sentence had ended. The occurrence of the clicks on the experimental tape triggered timing and storage of the responses on a micro-computer.

Results

Location task The results were analysed both for accuracy with respect to syllables and with respect to phonological segments. For example, for the syllabic analysis, each sentence was divided into syllables, as in:

| swing | of | the | pen | du | lum | was
<-... -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6... ->

I Responses before the objective click location were assigned a I negative value, responses after it a positive value. Marks through the vowel that contained the click counted as 0, marks in the same syllable as the click as 1, on the nearest syllable boundary as 2, etc. All responses further away than 2 syllables were assigned a 6. For the phoneme-by-phoneme analysis, the units were segment-sized, as on the response sheets. Since the two analyses of variance produced exactly the same pattern of results, we report the syllabic analysis only. Mean accuracy was -1.53 syllables away from the objective click location, replicating the previous finding that click migration is overwhelmingly leftward. There was no significant difference in the pattern of responses between English and Italian. There was a significant main effect of click location (Fl [1,46] = 11.45, $p < .002$; F2 [1,42] = 4.09, $p < .01$): listeners were more accurate in locating clicks in unstressed syllables than in stressed ones. There was also a main effect of word structure: listeners were more accurate in words that had primary stress on the second syllable. There was also an interaction of the above effects such that higher accuracy on clicks in unstressed syllables was found only for words which had primary stress on either their second *(analysis)* or their third *(metamorphosis)* syllable; the effect was the reverse, however, for words with primary stress on their first syllable *(lyricaf)*. This means that, overall, clicks were more accurately detected in earlier than in later position in the word. The apparent accuracy advantage of unstressed over stressed syllables is thus spurious; it arose merely because in two thirds of the word materials the unstressed syllables preceded the stressed. T-tests confirmed that the early-late difference was significant across both subjects and items for each word structure in each language.

Reaction time task Response times below 100 ms and above 1000 ms were discarded prior to analysis. The grand mean RT was 242 ms. No main effects or interactions were statistically significant across both subjects and items, though the factor of click location approached significance (F1 [1,46] = 9.87, $p < .01$; F2 [1,36] = 3.71, $p < .062$): RTs to clicks in stressed syllables were faster than those to clicks in unstressed ones. However, as in the previous task, there was an interaction of word structure with click location in the opposite direction for words with primary stress on their first syllables. So in fact, the apparent effect of stress on RT would also appear to be spurious: overall, RTs were faster to later occurring clicks. Thus, there was once more an effect of position of click, irrespective of language. Using a waveform editor, we measured durations of both the vowels in which the clicks had been placed and of the syllables these vowels were part of. Table 2 shows durations in milliseconds for the stressed and the unstressed vowels in both languages, with the duration of the unstressed vowels expressed as a proportion of that of the stressed vowels. As can be seen, the difference between stressed and unstressed vowels is smaller in Italian than in English. An analysis of variance on the proportions of the unstressed to the stressed vowel durations showed that the difference between the two languages was statistically significant (F [1,46] = 7.05, $p < .02$). We also conducted a correlation analysis of the measured vowel durations with reaction time. This showed that there was no statistically significant overall correlation of reaction time with vowel duration in either English or Italian, though there was a weak tendency for clicks in longer vowels to be detected faster in Italian.

Table 2. Duration (ms.) for stressed and unstressed vowels; % - duration of the unstressed vowels expressed as a proportion of the stressed ones.

	English			*Italian*		
	stressed	unstressed	%	stressed	unstressed	%
critical	58	29	50	60	46	75
metamorphosis	115	30	26	78	56	72
paralysis	106	25	24	134	67	50
retina	98	32	33	100	25	24
theoretical	83	31	38	146	85	58
electrical	102	66	64	124	123	99
lyrical	54	28	52	137	84	61
analysis	94	44	47	112	75	67
economical	80	12	15	91	60	65
formula	108	41	38	127	60	47
esoteric	95	42	44	123	59	48
*symme*trical	79	21	26	146	38	26
pendulum	75	41	54	105	70	66
philanthropist	83	16	19	126	54	43
monolingual	72	64	88	102	76	74
typical	62	55	87	57	56	97
political	74	11	14	92	49	53
catastrophic	94	36	38	113	53	46
synthesis	71	38	53	67	51	76
fantastic	90	64	71	73	52	72
parabolic	84	65	78	166	69	41
cynical	55	39	70	102	94	92
symptomatic	111	29	26	157	59	37
canonical	85	10	12	86	66	77

3. CONCLUSION

This investigation compared click detection performance in two languages that share the feature of lexical stress, but differ with respect to rhythmic characteristics as conveyed by syllable duration. The most noticeable feature of the results was that performance was essentially parallel in these two languages, replicating the Cutler *et al.* [13] findings for English and French. Moreover, in both languages and in both tasks, the strongest effect was exercised by the position of the click within the experimental item. In the location experiments, listeners were more accurate on clicks in syllables which occurred earlier in the word than on syllables which occurred later in the word. In the reaction time task, performance was better on clicks in later occurring syllables. Neither task showed any difference between Italian and English listeners' performance.

The processing of extraneous signals concurrent with speech input may well be sensitive to higher-level grouping effects [1,2,3,4]. Indeed, click detection has recently been used succesfully to study syntactic [14] and musical [15] parsing. However, given the existence of language-specific perceptual processing for English and Italian as demonstrated by Bertinetto and Fowler [12], the cross-linguistic similarity observed with the tasks used in this study suggests that click detection does not tap into language-specific processes of perceptual segmentation. We conclude that the process of click detection operates independendy from those levels of processing at which characteristics of language rhythm play a crucial role.

4. ACKNOWLEDGEMENTS

This research was supported by the ESPRIT BRA program [project P3207]. We thank Maddalena Agonigi, Lorenzo Ciono James McQueen and Costanza Papagno for their assistance.

5. REFERENCES

[1] *Fodor, J A.; Bever, T.G.:* The psychological reality of linguistic segments. J. Verb. Learn. Verb. Behav., Vol.4, pp. 414-420, 1965.

[2] *Fodor, J A.; Garrett, MJ[7].; Bever, T.G.:* Same syntactic determinants of sentential complexity, II: Verb structure. Perc. Psychoph., Vol. 3, pp. 453-461,1965.

[3] *Abrams, K.; Bever, T.G.:* Syntactic structure modifies attention during speech perception and recognition, Quart. J. Exp. Psych., Vol. 21, pp. 280-290, 1969.

[4] *Holmes, V.; Forster, K.:* Click location and syntactic structure. Perc. Psychoph., Vol. 12, pp. 9-15,1972.

[5] *Seitz,MR.;Weber,BA.:* Effects of response requirements on the location of clicks superimposed on sentences, Mem. Cog., Vol. 2, pp. 43-46,1974.

[6] *Ladefoged, P.; Broadbent, D.E.:* Perception of sequence in auditory events, Quart. J. Exp. Psych., Vol. 12, pp. 162-170,1960.

[7] *Titchener, E.:* Lectures on the experimental psychology of the thought processes. New York: McMillan, 1909.

[8] *Cutler, A.; Norris, D.:* The role of strong syllables in speech segmentation, J. Exp. Psych: Hum. Perc. Perf., Vol. 14, pp. 113-121,1988.

[9] *Cutler, A.; Butterfield, S.:* Rhythmic cues to speech segmentation: Evidence from juncture misperception. J. Mem. Lang., Vol. 31, pp. 218-236.

[10] *Mehler, J.; Dommergues, J.-Y.; Frauenfelder, U.H.; Segui, J.:* The syllable's role in speech segmentation, J. Verb. Learn.Verb. Behav.[7] Vol. 20, pp. 298-305,1981.

[11] *Vekds,D.;Bertinetto,P.M.:*Controllovs.compensazione: sui due tipi di isocronia. In: Magno Caldognetto, E.; Beninca, P. (Eds.) L'interfaccia tra fonologia e fonetica, Padova,pp. 155-162,1991.

[12] *Bertinetto, PM.; Fowler, CA.:* On sensitivity to durational modifications in Italian and English. Rivistadi Linguistica, Vol. 1, pp. 69-94,1989.

[13] *Cutler, A.; Kearns, R.; Norris, D.; Scott, D.:* Problems with click detection: insights from cross-linguistic comparisons, Speech Commun., in press.

[14] *Akeroyd, MA.;* Syntactic complexity effects in time-compressed speech. MPhil. Diss. Univ. Cambridge, 1992.

[15] Be*rent, I.;* *C.A.:* An on-line method in studying music parsing. Cognition, Vol. 46, pp. 203-222. 1993. I