On the Perception of Prominence in Short Phrases
A quantitative Experimental Study

Soundes Azabou - Kacem
University of Paris 7 – Denis Diderot
E-mail: Soundes.Azabou@linguist.jussieu.fr

ABSTRACT

This paper deals with two questions affecting the perception of prominence patterns in short phrases, by native English adults. In the first part, the influence of native linguistic knowledge on the perception of strongest prominence is evaluated for targets in sequences presenting potential stress clash. In part two, the influence of immediate target context on perceived prominence is studied. Three perception tests were made for this purpose. Perceived materials for each test were respectively filtered and excised targets, non-filtered clipped targets, and non-filtered targets in context. Results from test 1 vs. 2 show that it is easier for judges to assign perceived strongest prominence to "secondary" stress, when the segmental information is filtered, suggesting that auditory judgements for non-filtered targets might be affected by native linguistic knowledge. Results from tests 2 and 3 show a positive effect of context coupled with type of information and target position relative to nucleus.

1. INTRODUCTION

The aim behind this study is to investigate how the perception of strongest prominence operates in short phrases, namely the effect of epilinguistic knowledge, and the effect of following context on perceived strongest prominence. One way of testing the effect of native linguistic knowledge on judgements of prominence is making perception experiments involving non-native judges, who are not familiar at all with the language studied. While this method might provide acceptable results, the part attributed to the nature of the judge's mother tongue might not be always distinct. Another method for evaluating the effect of native knowledge can be achieved through stripping the stimuli off their lexical components, then reproducing the acoustic data and subsequently presenting the output to native adults. This method is used in part A. In part B, the effect of following context is studied using stimuli in various informational and intonational contexts, while syntactic and morpho-lexical information is kept constant. This is intended to keep the effect of these factors under control.

2. MATERIALS AND METHODS

2. 1. PRODUCTION

Perceived material consisted of six experimental sequences (ES) made up of the polysyllabic stress-shift word '*automatic*' (henceforth **target**), followed in some experimental sequences by an early accented noun '*pistol*'. The whole made up a sequence, '*an automatic pistol*', presenting a potential stress clash [1], and allowing for more than one prominence pattern to be produced and/or perceived. This basically depended on whether main stress in the target was shifted from the "primary" to the "secondary" syllable of the target or not, which was quite sensitive to the type of information (new or given) involved in the utterance. For this reason, the target had in every ES a different information type, and was pronounced in different informational contexts (cf. table1). Each ES covered a whole turn taking and was pronounced within one intonation unit. Seven female speakers of 'standard' British English produced the sequences in dialogue context (semi-spontaneous speech).

Number	Experimental Sequence	Information Type
ES1	*an* [*automatic*]$_N$	**New**
ES2	*an* [*automatic*]$_G$	**Given**
ES3	*an* [*automatic pistol*]$_N$	**New**
ES4	*an* [*automatic pistol*]$_G$	**Given**
ES5	*an* [*automatic*]$_N$ + [*pisto*l]$_G$	**New** + Given
ES6	*an* [*automatic*]$_G$ + [*pistol*]$_N$	**Given** + New

Table 1: Experimental sequence types as a function of type of information (N /New: information not presented in the discourse context, G /Given: information already presented in the context).

2. 2. PERCEPTION

Items for each perception test were digitalized in 16 bite AIFF sound files then randomised and edited within intervals of 8 seconds. For more accuracy in the responses, all items were repeated after a break of 2 seconds, longer silence intervals were respected before every new voice, and a muffled beep was inserted as a separation between items.

Thirty naïve adult English judges took part in the perception tests (table 2). After investigating the profile

ISBN 1-876346-48-5 © 2003 UAB

form filled in by the judges, seven among them were considered to be likely to have hearing problems. Their data were therefore excluded from analyses.

	Acoustic manipulation	Phrasal context
Test 1	Filtered target	Excised from context
Test 2	Non-filtered target	Excised from context
Test 3	Non-filtered target	In context

Table 2: Speech material in the three perception tests.

3. PART A: EFFECT OF EPILINGUISTIC KNOWLEDGE

The aim behind this experiment was to test the effect of native linguistic knowledge on perceived prominence. The question was whether prominence patterns in targets which had their lexical information filtered off (and therefore not recognized) were perceived similarly as patterns in non-filtered targets. For this purpose, test 1 and test 2 were designed.

3. 1. TEST 1: FILTERED TARGET

To hamper word recognition and therefore reduce to a minimum the effect of epilinguistic knowledge, targets from every ES were stripped of their lexical component. This was possible through filtering, after manipulation with *Praat* software. The obtained targets were then made up of three "hummed" speech segments that matched the original voiced sound sequences of the target "*a**n au**t(**o**)**ma**t**i**c*"(shown here in underlined bold). The first 'hummed' segment represented the filtered 'secondary' syllable /nɔː/, the second segment included the filtered 'primary' syllable /mæ/ and the third segment was the filtered vowel of the last syllable of the target /tɪk/.

The judges task was to tick the box corresponding to the most prominent speech segment among the first two, or tick the box '*unclear*' if locating the strongest 'hummed' prominence was not possible.

3. 2. TEST 2: NON-FILTERED TARGET

In test 2, the same judges, who took part in test 1, heard the non-filtered target excised from its phrasal context. Their task was to locate the strongest syllabic prominence in the target, i.e. /nɔː/ or /mæ/. In case both syllables were perceived as "really ***equally*** prominent", there was an extra choice: ticking the box corresponding to '*both syllables*'.

3. 3. RESULTS: TEST 1 vs. TEST 2

A statistic *Square Chi* test was performed on the data to test whether there was similarity in the distribution of perception values between filtered (test 1) and non-filtered targets (test 2). The hypothesis was that value distribution of filtered items was similar to that of non-filtered targets. Now, the threshold value for significance being (p=0,05), results showed that the hypothesis was null and void (p = 0,03 ; *Square Chi* = 136 ; df = 107), which means that the value distribution for filtered targets and that for non-filtered targets were *significantly differential*. Such dissimilarity may have been caused by the absence of the lexical component in 'hummed' targets.

Now, the point was to check in what way the distribution was dissimilar. The prediction was that perceived strongest prominence should be assigned *more easily* to 'secondary' syllable /nɔː/, in filtered items than in non-filtered items. The reason is that in 'hummed' targets, if acoustic cues and parameters are not sufficient to signal strongest prominence on the speech segment matching /mæ/, there is no access to judgements *via* correction, since there is no access to word recognition. Perception results from test 1 and test 2 showed, indeed, that in all types of experimental sequences, perception values for strongest prominence on initial syllable/ speech segment were higher in filtered than in non-filtered targets (cf. figure 1). It seems easier, then, for judges to assign perceived strongest prominence on the 'hummed' speech segment matching the secondary syllable (43% of judgements) than on the non-filtered secondary syllable itself (32%). Such inconsistency in the assignment of prominence is certainly due to the lack of the lexical information in filtered targets, and therefore signals the inaccessibility to word recognition and to epilinguistic judgements.

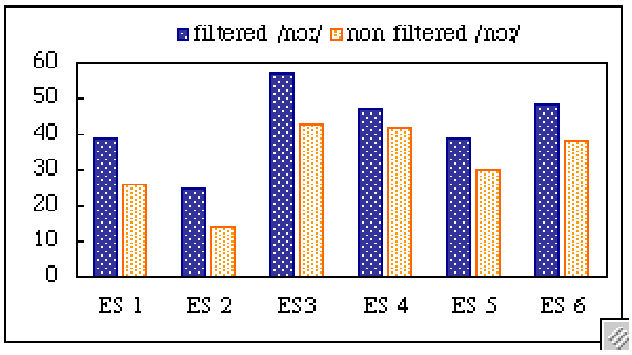

Figure 1: Perception of strongest prominence at secondary syllable /nɔː/ in filtered and non-filtered target: '*an automatic*' /ə nɔː təmæ tɪk/, as a function of sequence type. *Y - axis*: Judgements of prominence in the target (%). *X - axis*: Sequence type as a function of type of information in the phrase (N: New, G: Given). For details of sequence type (ES), cf. table 1.

3. 4. DISCUSSION

Comparative results from test 1 and test 2 showed dissimilarity in perceived prominence. This was attributed to the absence of the lexical component in filtered targets. Other minor acoustic and/or methodological constraints might have contributed, as well, to the variability, namely the difference in the options presented to the judges in each test, i.e. the option "unclear" instead of "both" (both segments are *equally* prominent) in filtered target (table 3). This is due to the substantial difficulty of evaluating consecutive 'hummed' sounds. Indeed, up to 80 % of the

judges (who were asked) thought it was quite difficult to describe 'hummed' sounds.

Filtered target			*Non-filtered target*		
1st segment	2nd segment	Unclear	/nɔː/	/mæ/	Both
43 %	40 %	18 %	32 %	37 %	30 %

Table 3: Mean values of perceived strongest prominence in filtered and non-filtered targets.

4. PART B: EFFECT OF POST-TARGET CONTEXT

The second part of this paper dealt with the effect of hearing the post-target context on judgements of perceived strongest prominence. The pre-target context was, therefore, made constant in all experimental sequences (the same article *an* at beginning of turn taking).

4. 1. TEST 3: TARGET IN CONTEXT

In test 3, the perceived items consisted of the target, as naturally produced within its original context, that is the entire noun phrase '*an automatic pistol*' as produced in sequences ES3 to ES6 (table1). The judges' task, as in test 2, was to locate the strongest syllabic prominence.

4. 2. TEST 2 vs. TEST3

4. 2. 1. New targets

A *Square Chi* test was performed on perception data from test 2 and 3, for new targets (ES3: new pre-nuclear targets, and ES5: new nuclear targets) to check if there was similarity in the distribution of variables in context (test 3) and out of context (test 2). Results showed a *significant dissimilarity* in distribution between both sets of variables ($p = 0{,}001$; *Square Chi* = 74 ; $df = 41$). In effect, values for perceived strongest prominence for each target are more clear-cut in test 3 (cf. figure 2). A ratio to calculate this distinction (*D. ratio*) is obtained by dividing perception values for syllable /nɔː/ by values for /mæ/.

For new pre-nuclear targets (ES3), the value for *D. ratio* was 1,3 for targets out of context, while in context it went up to 3. The same tendency was confirmed in new nuclear targets (ES5) where the value for *D. ratio* doubled for targets in context (1,5 vs. 3,6).

4. 2. 1. 1. Discussion

Such inconsistency in the assignment of prominence highlights the positive effect of post-target context. It seems, in fact, easier for judges to assign perceived prominence when the target is perceived within context. In context, judges have access to more elements and cues to interpret the prosodic and intonational choices taken by the speaker. For instance, in ES3 the absence of *Low Fall*, in most targets is better interpreted in context rather than in targets excised from context (where the isolated target is expected, as in citation form, to be nuclear and probably to end in a *Low Fall* contour).

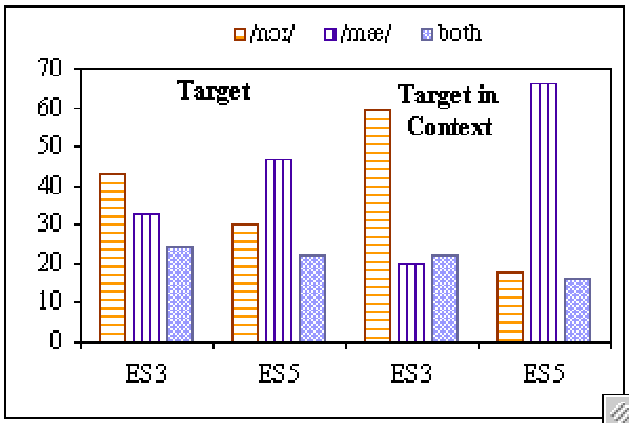

Figure 2: Distribution of perceived strongest prominence in 'new' targets (*an automatic*) as a function of context.
Y axis: Mean judgements of strongest prominence (%).
X axis: Sequence with 'new' targets, as a function of context.

4. 2. 2. Given targets

Distributions for given pre-nuclear targets (**ES4**: 'given' target plus 'given' noun ; and **ES6**: 'given' target plus 'new' noun), in test 2 and test 3, seem to be different (figure 3). Again, the Distinction ratio is used to evaluate difference. Indeed, the more the *D. ratio* between syllables is close to 1, the more syllables are likely to be perceived as similarly prominent. In test 2, that is when the target is clipped out of following context, the value of *D. ratio* between syllables /nɔː/ and /mæ/ is 2 in SE4 (significant difference, $p = 0{,}01$) and 1,6 in SE6. Yet, in context, *D. ratio* falls down respectively to 1,1 (non-significant difference, $p > 0{,}05$) and 1,2. This means that distributions of perception values for /nɔː/ and /mæ/ are distinct out of context and similar when the given target is perceived within context.

According to 40 % of judges, it is the "secondary" syllable /nɔː/ that bears the strongest prominence out of context. In context, it is rather the option 'both' that has the highest distribution for strongest prominence, with 41 % of prominence judgements (cf. figure 3).

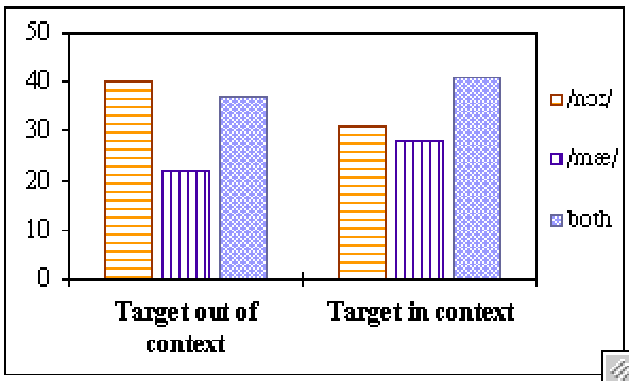

Figure 3: Distribution of perceived strongest prominence in 'given' targets (*an automatic*) as a function of context.
Y *axis*: Mean judgements of strongest prominence (%).
X axis: Sequences with given targets, as a function of context.

ISBN 1-876346-48-5 © 2003 UAB

4. 2. 2. 1. Discussion

Now the question is why perception of strongest prominence equally on both syllables is favoured when given targets are perceived in context. This is certainly due to the fact that clipped given targets might be considered by judges as nuclear, for they are isolated, whereas in context, given targets seem to be clearly perceived as pre-nuclear. In other words, the presence of a nucleus in the context of a given target helps the judges to interpret the target as relatively toned down or even de-accented, hence the judgement of "primary" and "secondary" syllables as equally prominent. However, it is worth noting that, at variance with [2], not all pre-nuclear targets can be perceived as having a levelled prominence pattern for, in our material, this only concerns **given** pre-nuclear targets. Indeed, in ES3, new pre-nuclear targets were perceived with strongest prominence on the "secondary" syllable.

The results presented above reveal that more than one syllable is perceived as bearing the strongest prominence in the target, which seems contradictory to the treatment of prominence as being an absolutely hierarchical property [1].

5. CONCLUSION

In part A, the incidence of native linguistic knowledge on perceived strongest prominence was demonstrated. When reference to epilinguistic knowledge was hampered through filtering, it was easier for judges not to stick to canonical prominence patterns. Part B, provided evidence for the role of following context in prominence perception. For new targets, results were more distinct and probably more accurate in context. For given pre-nuclear targets, the role of context was even more conspicuous, since hierarchical distinctions were blurred in the context of a following nucleus.

In short, results gave more clues about the nature of perceived prominence. There is evidence that judgements of prominence do not necessarily constitute the simple equivalents of their acoustic variables. Native linguistic knowledge, informational and intonational context, and possibly other factors like usage frequency or degree of lexicalisation could substantially affect prominence judgements in short phrases. Further experiments are therefore necessary to determine the various factors affecting prominence perception in short and longer phrases.

REFERENCES

[1] M. Liberman and A. Prince, "On stress and linguistic rhythm," *Linguistic Inquiry*, vol. 8, pp. 249–336, 1977.

[2] E. Grabe and P. Warren, "Stress shift: do speakers do it or do listeners hear it?", in *Papers in Laboratory Phonology IV*, B. Connell and A. Arvaniti, Ed., pp. 95–110. Cambridge University Press, 1995.