Processing focus and accent across dialects.

Sabine Zerbian, Giuseppina Turco

PROCESSING FOCUS AND ACCENT ACROSS DIALECTS[i]

Sabine Zerbian & Giuseppina Turco

University of Stuttgart, Department of English Linguistics
Laboratoire de Linguistique Formelle (UMR 7110), CNRS-Université Paris Diderot, Paris
sabine.zerbian@ifla.uni-stuttgart.de - gturco@linguist.univ-paris-diderot.fr

ABSTRACT

Previous research has shown that native listeners of British English integrate semantic and prosodic cues such that the presence of both sources of information does not facilitate phoneme detection. In the study reported here, we show that listeners of General South African English listening to British English material show a different pattern, using both prosodic and semantic information in phoneme detection. As previous research has shown that listening to regionally-accented speech affects processing at different levels, we attribute this difference to the fact that these listeners are listening to a non-native dialect.

Non-native listeners (e.g. Dutch listeners listening to English material) show a similar processing mechanism in using prosodic and semantic information independently. It thus seems that processing mechanisms are adapted not only when listening to a foreign language (even for highly proficient speakers of a foreign language) but also when listening to a different dialect of the same language.

Keywords: dialect, processing, British English, South African English, prosody, phoneme detection

1. INTRODUCTION

In English, the information structure of a given utterance can be marked prosodically, e.g. in that focused constituents are realized with a pitch accent and given constituents are deaccented [9]. The use of prosody used for marking information structure has been reported for different varieties of English (here referred to as dialects), such as British English, American English and General South African English to name a few. Similarly, also other languages, such as Dutch and German, mark information structure in a similar way.

For speech processing, previous research has shown that English listeners exploit both semantic information (i.e. focus, [5]) and prosodic cues (i.e. accent, [3]). Furthermore, [1] showed that semantic and prosodic cues are not processed independently of each other when manipulated in the same experiment: "when listeners are given semantic cues as to where to find the new information in an utterance, the search for accent has less to offer" [1:86].

The current study seeks to explore whether the integrated processing of prosodic and semantic cues also holds when listening to a different dialect of one's first language. [10] researched the effects of regionally-accented speech at different processing levels and found that phonemic discrimination and word recognition are most affected, whereas semantic priming does not seem to be affected much by listening to dialects. [7] has shown that listening to regionally-accented speech has costs on spoken word recognition. These costs can be long-lasting with no habituation after repeated exposure to the same accent ([6]).

By means of a phoneme detection task, the current study investigates the processing of semantic focus and its prosodic cues in white monolingual listeners of General South African English, listening to British English stimuli. General South African English (GenSAfE) and British English (BrE) share phonological characteristics due to their shared history, such as being non-rhotic ([2]). Both varieties are also very similar in their prosody by manipulating pitch, duration and intensity for semantic focus marking ([12]). At the same time, BrE and GenSAfE are two clearly discernible regional varieties of English, with the two main phonological indicators being the vowels of KIT and BATH ([2]). In addition, a linguistic orientation towards American English can be observed in young South African speakers, possibly due to media influence (see [8]). This suggests that BrE is perceived as a marked accent in South Africa.

Our prediction is that processing of regionally-accented speech will incur overall higher processing costs which might not underlie habituation even after extended exposure to the accent. In a phoneme-detection task, this is expected to result in longer reaction times when listening to regionally-accented stimuli as compared to same-dialect stimuli. Next to quantitative differences, there might also be qualitative differences. It is an open question whether the linguistic similarities between these two varieties of English lead listeners to integrate prosodic and semantic cues in a similar way that listeners listening to the same dialect will, or whether a different processing strategy is used when listening to regionally-accented speech.

2. PROCESSING FOCUS AND ACCENT

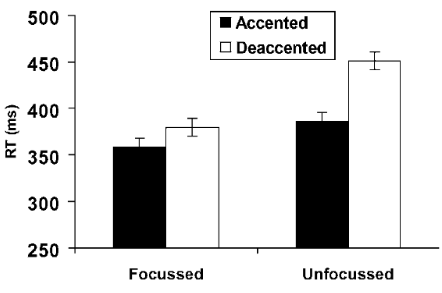
By means of a phoneme-detection task, [1] tested English and Dutch speakers on their exploitation of semantic and prosodic cues in speech processing. Participants had to press a button as soon as they heard a pre-specified sound. The set of materials comprised twenty-four semantically unrelated experimental sentences, like for instance “the young man on the corner was wearing a blue hat”, each containing a pre-specified target phoneme (either /d/ or /k/ or /b/; e.g. /k/ of *corner*).

Semantic status was controlled for by means of a preceding question, which asked either for the target word containing the pre-specified phoneme (focused) or some other constituent (unfocused). For example, with [b] being the target phoneme in the sentence “The man at the corner was wearing the blue hat.”, the question “Which hat was the man wearing?” would yield the focused condition, whereas the question “Which man was wearing the hat?” would yield the unfocused condition.

Prosodic status was controlled for by having recorded the target sentences in different focus conditions, with accent on either the target word or some other word. Although the target-phoneme bearing word itself never bore an accent in the experimental stimuli (having been cut and spliced in from a neutral reading), the stretch preceding the target word showed differences in prosody.

The results for British English listeners are reproduced in figure 1.

Figure 1: Mean response times (ms) for the effect of predicted accent as a function of focus (taken from [1: 86])



Results show that the target phonemes are generally detected faster in focused words than in unfocused words. Only in unfocused words is there an additional benefit of accent, but not in focused words. The same pattern emerged for Dutch listeners listening to Dutch, which has a similar focus-to-accent-mapping (cf. [1: 87, fig. 2]). As stated in the introduction, findings were interpreted in that prosodic and semantic cues are not exploited together for the same processing purpose when both are available. If both cues match, processing is fastest, though not significantly faster than if only the semantic cue is present. If both cues are absent, however, processing is considerably delayed.

3. METHODOLOGY

We replicated [1]’s phoneme detection study, using the original BrE stimuli. The participants’ task was to click a button as soon as they heard a pre-specified sound. The reaction times to the button presses were measured and interpreted as cues to processing.

3.1. Materials

The materials were those ones used in [1]. The set of materials comprised twenty-four semantically unrelated experimental sentences, each containing one target phoneme (either /d/ or /k/ or /b/) in the early position of the sentence (e.g. /k/ of *corner*) and another one in the late position of the sentence (e.g. /b/ of *blue*).

Semantic status was controlled for by means of a preceding question, which asked either for the target word containing the pre-specified phoneme (focused) or some other constituent (unfocused).

Prosodic status was controlled for by having recorded the target sentences in different focus conditions, with accent on either the target word or some other word.

The different manipulations resulted in eight versions of each target sentence, with all possible combinations of early/late target, focused/unfocused and two prosodic contexts (accented/unaccented). Each participant heard only one of these eight versions. Hence, eight different lists were created to which participants were randomly assigned. In addition, the experimental material contained 24 filler sentences.

3.2. Participants

Forty-nine monolingual white speakers of GenSAfE between the age of 19 and 29 (25 male, 24 female, age average= 21.5, SD=4.4) took part in the study. All of them had (GenSAf) English as their only first language. Participants were all students at the University of Witwatersrand, Johannesburg. They received a travel reimbursement fee for their participation. None of them reported vision or hearing impairments.

3.3. Procedure

The testing procedure was a faithful replication of the one by [1] in terms of instructions, order of the stimuli presentation and inter-stimulus interval (2 seconds). The original recordings were used[ii], which were produced in a unmistakenly British English accent. In each trial, participants first saw the target phoneme (appearing for one second on the screen of a portable laptop) and then heard (binaurally over headphones) the question followed by the answer. After the phoneme-detection, a post-recognition test was administered in a pen-and-paper version to test actual language comprehension during the phoneme-detection task. It was a multiple choice test comprising 24 of the 48 target sentences heard during the phoneme-detection task. Subjects had to decide between four words, which had been the (early or late) target-bearing word.

The experiment was designed by using E-prime software [11]. Response times were recorded by means of a button box linked to the portable computer and calculated in relation to the timing interval between the start of the sentence and the onset of the target-bearing word. Participants were tested individually in a quiet room at the University of the Witwatersrand, Johannesburg.

4. RESULTS

4.1. Post-recognition test

The overall mean score of correct answers to the post-recognition test was 75% (SD = 12.5).

4.2. Phoneme-detection test

No subject was responsible of more than 3 missing responses. Following [1], one item was excluded from the analysis to avoid confounding factors in the analysis (see sentence 19 in Appendix in [1]).

Concerning the accuracy data (number of correct responses in the detection of the phoneme), table 1 shows the number of missing responses (43 overall, coded as "N(o)") split by Prosodic status (A: accented, UNA: unaccented) and semantic status (F: focused, UNF: unfocussed).

	A		**UNA**	
	F	**UNF**	**F**	**UNF**
N	8	9	10	16
Y	274	273	272	265
	282	282	282	281[*iii]

Table 1: Missing responses ("N(o)") and correct responses ("Y(es)") in phoneme-detection task

Statistical significance was checked with a binomial logistic regression model with type of response (correct "1" vs. incorrect "0") as dependent variable, semantic status (focused vs. unfocused) and prosodic status (accented vs. unaccented) as predictors. Random slopes for each factor were included in speaker and item. There was no effect of semantic status (p=0.4), no effect of prosodic status (p=0.6) and no interaction (p=0.2).

As for reaction times (RTs), missing responses (43 data points) were excluded, thereby leaving 1084 data points for analysis. Furthermore, since responses less than 100 ms are more likely to reflect guessing (resulting from anticipation) of the target-bearing word, while RTs longer than 1500 ms might be interpreted as a sign of reprocessing the whole sentence (see [1, 3]), RTs shorter than 100 ms and longer than 1500 ms were also excluded (N=52, accounting for 4.8% of the data). This left 1032 responses for the RTs analysis.

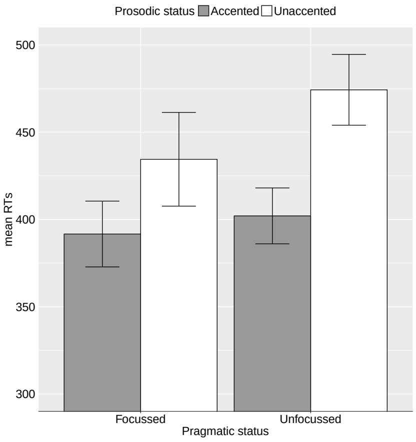
The overall RTs mean (423.4 ms) were slightly higher, and thus slower, than that of the BrE group (mean: 394 ms) tested in [1]. Furthermore, a mean difference of 33 ms was found between RTs for early targets vs. RTs for late targets (early: 439.6 ms vs. late: 406.2 ms), in line with previous studies showing that early targets are generally detected slower than late ones, [1].

We run a linear mixed effects model (based on 49 speakers, 23 items and 1032 observations) with RTs as function of semantic status (focused vs. unfocused), prosodic status (accented vs. unaccented) and target position (early vs. late). The model revealed:

- A main effect of semantic status: Target phonemes in focused words were detected significantly faster than in unfocused ones ($\beta_{focused}$= 24.3, SD= 9.3, t-value= 2.59, p< .05)
- A main effect of prosodic status: Target phonemes in accented words were detected significantly faster than in unaccented ones ($\beta_{accented}$= 55.5, SD= 9.9, t-value= 5.58, p< .0001).
- A main effect of target position: Late target phonemes were detected significantly faster than early ones (β_{late}= 35.7, SD= 14.4, t-value= 2.48, p< .05)
- No interaction between any of the three factors (all p-values> 0.1).

Figure 2 shows the mean RTs for the effects of prosodic and semantic status (focused/accented: 392.4 ms; focused/unaccented: 434.3 ms; unfocused/ accented: 401.8 ms; unfocused/unaccented: 470.5 ms).

Figure 2: Mean response times (ms) (GenSAfE listeners tested on BrE materials), whiskers represent standard errors



An ANOVA analysis following the procedure as in [1] confirms the results of the linear mixed effects model.

5. DISCUSSION

Listeners of GenSAfE exploit both semantic and prosodic information independently in the processing of a BrE sentence, as evidenced by the two main effects of prosodic status and semantic status. Accents lead to significantly faster reaction times both in focused and unfocused words. It is in this respect that processing of a regionally-accented English differs from same-dialect processing. In same-dialect processing (BrE listeners listening to BrE material in [1], fig. 1) an interaction between focus and accent was found, such that accent contributes only little in focused contexts.

For across-dialect processing, we postulated two hypotheses, namely that overall reaction times are slower and that qualitative differences emerge in the processing of semantics and prosody. As for the quantitative differences, the overall mean RTs are considerably longer in the across-dialect study presented here (namely 423 ms) than in the same-dialect study conducted by [1] (394 ms). However, for various reasons it is not valid to directly compare the absolute values across [1] and our study, so this is left for future research.

On the other hand, a qualitative difference in the processing of semantic and prosodic cues within and across dialects clearly emerges. In same-dialect processing (see figure 1) semantics and prosody are exploited with no facilitatory effect if both match, but considerable delay if neither prosody nor semantics is guiding listeners' expectations. In across-dialect processing (see figure 2) both semantics and prosody are exploited in processing. Prosody always has a facilitatory effect independent of semantic cues. Thus, there is an integrated processing mechanism for prosodic and semantic cues in same-dialect processing, not influencing each other to any significant extent if both cues converge. In across-dialect processing, however, processing of prosodic and semantic cues remains independent.

[1] use the term „fail-safe, belt-and-braces approach" for a processing strategy that does not integrate both kinds of information but use the cues independently of each other, contrary to same-accent L1 listeners. [1] observe this processing strategy also in non-native listening, when Dutch listeners listened to English material. Despite similarities of English and Dutch concerning the use of prosody to mark focus, Dutch listeners who are highly proficient in English show a processing strategy in non-native listening that is different to the processing strategy they show when listening to Dutch material.

Dutch listeners listening to BrE material show main effects of prosodic and semantic status but no interaction (cf. [1: fig. 4]). Thus, they too, use prosody and semantics independently in processing. The parallel processing pattern can also be found South African Sepedi-English bilinguals listening to BrE material (Turco & Zerbian, under review).

The results of our study therefore show a qualitative difference in phoneme detection across and within dialects. Whereas within the same dialect, prosodic and semantic cues are integrated, they are processed independently across dialects. As a parallel observation has been made for non-native listening, it seems that bilingual and across-dialect listening resemble each other in this particular aspect. Native English listeners listening to a variety of English other than their own thus seem to pattern with listeners listening to a different language rather than English L1 listeners listening to their own variety in this particular aspect.

6. REFERENCES

[1] Akker, E., Cutler, A. 2003. Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition* 6(2), 81–96.

[2] Bowerman, S. 2008. White South African English: phonology. In: Mesthrie, R. (ed), *Varieties of English 4 – Africa, South and Southeast Asia*. Berlin/New York: Mouton de Gruyter, 164–176.

[3] Cutler, A. 1976. Phoneme monitoring and intonation contour. *Attention, Perception, & Psychophysics* 20(1), 55–60.

[4] Cutler, A. 2012. *Native Listening*. Cambridge, Mass.: The MIT Press.

[5] Cutler, A., Fodor, J.A. 1979. Semantic focus and sentence comprehension. *Cognition* 7, 49-59.

[6] Floccia, C., Butler, J., Goslin, J., Ellis, L. 2009. Regional and foreign accent processing in English: can

listeners adapt? *J Psycholinguistic Research* 38(4), 379-412.
[7] Floccia, C., Goslin, J., Girard, F., Konopczynski, G. 2006. Does a regional accent perturb speech processing? *Journal of experimental psychology. Human Perception and Performance* 32(5), 1276-93.
[8] Hartmann, D., Zerbian, S. 2009. Rhoticity in Black South African English- A sociolinguistic study. *Southern African Linguistics and Applied Language Studies* 27(2), 135-148.
[9] Ladd, D.R. 1996. *Intonational phonology*. Cambridge University Press.
[10] Larraza, S., Samuel, A.G., Oñederra, M.L. 2017. Where do dialectal effects on speech processing come from? Evidence from a cross-dialect investigation. *The Quarterly Journal of Experimental Psychology* 70(1), 92-108.
[11] Psychology Software Tools, Inc. 2017. *E-Prime: Documentation Article*. Retrieved from https://support.pstnet.com.
[12] Zerbian, S. 2015. Prosodic marking of focus in transitive sentences in varieties of South African English. In: Gut, U., Fuchs, R., Wunder, E. (eds) *Universal or diverse paths to English Phonology*. Berlin: De Gruyter, 209-240.

---

i This research was funded by the German Research Foundation (DFG), grant to the SFB 732, project A7. We are grateful for the assistance of Khutso Matlou in preparing and conducting the experiment as well as for the support of the Department of Linguistics at the University of the Witwatersrand, Johannesburg.

ii We are grateful to Anne Cutler for allowing us to use and providing us with the original recordings.

iii The sum of the last condition is 281 instead of 282 because the number of items is 23 (instead of 24), which cannot be divided by 4.