TIME ORDER ERROR IN SPEECH PERCEPTION: ELECTROPHYSIOLOGICAL EVIDENCE

Hans Georg Piroth* and Friederike Schlaghecken†

**Institute of Phonetics and Speech Communication, Ludwig-Maximilians-University Munich,Germany,*

†Department of Experimental Psychology, Cambridge University, England

ABSTRACT

To investigate how temporal features of speech sounds are processed, electrophysiological data (AEP) were registered in a deviant detection task. Stimuli were presented in an oddball design, employing a 5-step voice onset time (VOT) continuum from [ba] to [pa]. Deviant AEPs showed both an enhanced N1- and N2b component compared with standard AEPs. Moreover, when standards had a longer VOT than deviants, deviant AEPs showed a substantial reduction of the P2 component. This effect was missing when the same stimuli were presented in a context of standards having a shorter VOT. It appears that this effect is due to an earlier onset of the N2b component in the former condition, indicating that conscious detection of deviant stimuli is easier under conditions where a short VOT is matched against a template of a long VOT.

1. INTRODUCTION

Within the framework of a project aimed at gathering information on autonomous reactions to the process of speech perception from multiple data deviation sources, electrophysiological data (AEP) were registered to investigate how temporal features of speech sounds are processed. As known from the literature on AEPs in general, acoustic stimuli evoke a series of specific negative- and positive-going mid latency deflections: A negative peak (N1-component) occurs around 100 ms after stimulus onset, followed by a positive peak (P2-component) around 180 ms. If the hearers' attention is directed to the stimuli, and if a stimulus deviates from what the perceiver has expected to hear, a second negativity (N2b-component) is elicited, starting around 200 ms after stimulus onset. If AEPs elicited by deviant (unexpected) stimuli are compared with AEPs elicited by standard (expected) stimuli, the former are found to be more negative-going in an even earlier latency range (about 150 ms to 200 ms after stimulus onset; mismatch-negativity: MMN) [1]. If the stimuli are task-relvant, a positive-going peak (P3-component) develops about 300 ms after stimulus onset [2].

After some initial AEP studies on speech perception in the seventies [3,4], an increasing number of studies have used the categorical perception paradigm to detect electrophysiological correlates of speech sound categorization. Phonetic "continua" varying acoustic signal parameters stepwise were synthesized to yield a stimulus set "continuously" moving from one clearly perceivable phoneme category to another [5,6,7,8]. However, these studies were restricted almost completely to an investigation of MMN effects, mainly addressing the question whether the MMN has to be regarded as reflecting a linguistic or a pre-linguistic level of processing.

The present experiment was planned to re-inspect mid latency components by use of a simple temporal acoustic variation throughout a phonetic continuum in order to account for the divergence of the results obtained in the earlier experiments. Special attention was given to the fact that the order of stimulus presentation is critical for discrimination tasks. As known from literature [9,10] - and confirmed for different modalities [11] as well as for behavioral data on VOT [12] and intonation [13] in speech - performance in discrimination of stimulus pairs is better when a more intensive or shorter stimulus follows a less intensive or longer one ("time order error": TOE). Since EP recordings are usually collapsed across stimulus order for analysis, this effect might be covered in the description of several other studies.

2. METHOD

A 5-step voice onset time (VOT) continuum from [ba] to [pa] was synthesized by means of the Klatt speech synthesizer [14] varying VOT in equal steps from 10 to 50 ms, but keeping other parameters contributing to the voicing distinction constant at intermediate values between [ba] and [pa]. (All stimuli started with a 5 ms burst (64 dB) without any aspiration. F0 start value was set to 120 Hz dropping to 101 Hz in the final part. F1 transition started at 300 Hz to 620 Hz steady state, F2 and F3 were accordingly set: 1100 Hz to 1220 Hz and 2080 Hz to 2250 Hz, respectively.)

9 volunteers (5 male), aged 27 - 60 years (mean age: 34.7 years) participated in the experiment. All participants but one were right-handed and had normal or corrected-to-normal vision. Participants were seated in a dimly lit, sound attenuated and electrically shielded chamber, with a response button under their right index finger. Stimuli were delivered via headphones.

The experiment consisted of 20 blocks, delivered in two sessions of 10 blocks each. In each block, one stimulus was presented as 'standard stimulus' on 425 trials (85% of all trials), a second stimulus was presented as 'deviant stimulus' on 75 trials (15%), resulting in a total of 500 trials per block. At the beginning of each block, at least 10 standard stimuli were presented. Since all stimulus combinations were tested, each of the five stimuli served in four blocks as standard stimulus, and in another four blocks as deviant. Blocks were delivered in such a sequence that neither the standard nor the deviant stimulus of a given block appeared as standard or as deviant in the next block. The interstimulus interval (ISI) was randomized within each block and was either 550 ms, 600 ms, or 650 ms, each ISI appearing with equal probability. Participants were informed that they would hear the standard stimulus about 10 times at the beginning of each block and that, some time after this initial sequence, they could expect to hear a deviant stimulus on several trials throughout the block. They were instructed to respond with a button-press in case they had noticed the occurrance of a deviant. At the end of each block, participants received feedback about the number of hits and false alarms.

EEG was recorded with Ag-AgCl electrodes from Fpz, Fz, Cz and Pz (according to the international 10-20 system) and from the left and right mastoid. All electrodes were referenced against the nose. Impedance was kept below 5 kOhm. EEG was sampled with a digitization rate of 256 Hz, amplifier bandpass was 0.1-70 Hz. EEG was averaged off-line for epochs of 600 ms, starting 100 ms before stimulus onset, and ending 500 ms afterwards. Epochs containing eyeblinks (FPz-voltage exceeding +/- 50 µV) were excluded from analysis. Averages were computed separately for standards and deviants for each block, resulting in a total of 40 AEP waveforms per participant. To test oddball effects, AEPs elicited by physically identical stimuli were compared, e.g., AEPs elicited by VOT10 standard stimuli were compared with AEPs elicited by VOT10 deviant stimuli. To investigate VOT effects, all AEPs elicited by a given stimulus were grouped together, separately for their occurance as standard and as deviant stimuli, resulting in 10 AEP waveforms for each participant (deviant VOT10, deviant VOT20, ... ; standard VOT10, standard VOT20, ...). To test step-size effects, AEPs were collapsed separately across the eight 1-step blocks, across the 6 2-step blocks, across the 4 3-step blocks, and across the 2 4-step blocks, respectively, again separately for standards and deviants, resulting in 8 AEP waveforms for each participant. To test serial order effects, all AEPs elicited by stimuli that occurred together with stimuli of shorter VOTs were grouped together, as well as all AEPs elicited by stimuli that occured together with stimuli of longer VOTs, again separately for standards and deviants, resulting in 4 AEP waveforms for each participant (deviant longer VOT, standard longer VOT, deviant shorter VOT, and standard shorter VOT).

After visual inspection of the grand mean average waveforms, the N1 latency range was determined as the 75 ms - 125 ms time window after stimulus onset, the P2 latency range as the 150 ms - 250 ms time window, and the N2b latency range as the 200 ms - 350 ms time window. Statistical analysis was performed on the peak amplitudes values in the N1 and N2b latency range, and on the mean amplitude values in the P2 latency range. For oddball effects and serial order effects, repeated measures ANOVAs were performed for the factors TYPE (standard, deviant), ORDER (shorter vs. longer VOT), and ELECTRODE (Fpz, Fz, Cz). For VOT effects, repeated measures ANOVAs were performed for the factors VOT (VOT10 to VOT50), TYPE, and ELECTRODE. For step-size effect, repeated measures ANOVAs were performed for the factor STEPSIZE (1-step, 2-step, 3-step, 4-step), TYPE, and ELECTRODE. Where appropriate, Greenhouse-Geisser adjustments to the degrees of freedom were performed (indicated in the Results section by ‘ε’). Results from mastoid electrodes are not reported in this paper. Hits (correctly detected deviants) and false alarms (standards incorrectly classified as deviants) were recorded for each block. Statistical analyses were performed on hit rates and false alarm rates similar to the EEG analyses.

3. RESULTS

Oddball effects: As can be seen in Fig. 1, deviant stimuli elicited larger negative-going deflections than standard stimuli in both the N1 latency range ($F(1,8) = 14.43$, $p = .005$) and in the N2b latency range ($F(1,8) = 8.74$, $p = .018$). In the P2 latency range, the main effect of TYPE only approached statistical significance ($F(1,8) = 4.62$, $p = .064$).

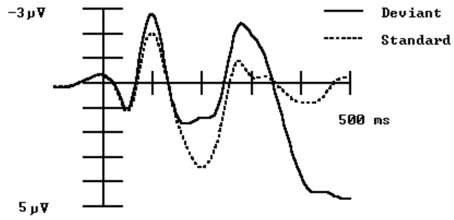

Figure 1. Effects of stimulus deviance on AEP waveforms: Grand mean average waveforms at Cz, collapsed across all stimulus types (VOT10 - VOT50), plotted separately for deviant items (solid line) and standard items (dashed line).

There was a VOT main effect on both hit rate ($F(4,32) = 19.73$, $p < .001$, $\varepsilon = .551$) and false alarm rate ($F(4,32) = 5.24$, $p = .015$, $\varepsilon = .547$), as hit rate was higher and false alarm rate was lower for stimuli from the egdes of the continuum than for stimuli from the center (see Fig. 2). VOT also influenced N1-amplitude ($F(4,32) = 5.01$, $p = .027$, $\varepsilon = .433$), as the N1 peak increased with decreasing VOT (see Fig. 3).

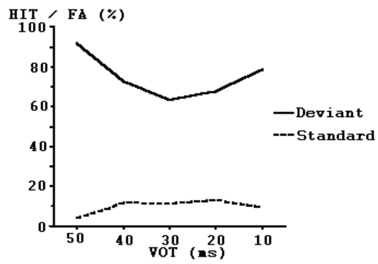

Figure 2. Effects of VOT on detection performance: Hit rate (solid line) and False alarm rate (dashed line) for each of the five stimulus types, collapsed across all pairings.

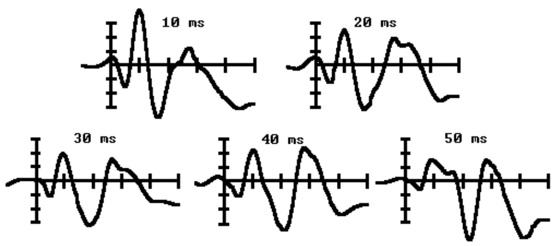

Figure 3**.** Effects of VOT on AEP waveforms: Grand mean average waveforms at Cz, plotted separately for each of the five stimulus types, collapsed across all pairings.

With increasing step-size, hit rate increased from about 50% to almost 100%, while false alarm rate dropped from about 20% to almost zero (see Fig. 4), resulting in a highly significant effect of STEPSIZE on both hit rate and false alarm rate (F(3,24) = 73.53, p < .001, ε = .567; and F(3,42) = 20.12, p = .001, ε = .443, respectively). Correspondingly, N1-amplitude increased with increasing step-size (Fig. 5) for both standards and deviants (F(3,24) = 4.16, p = 0.41, ε = .598). Although Fig. 5 seems to suggest that step-size also influenced P2 and N2b-amplitude, these effects failed to reach statistical significance (all F < 2.6).

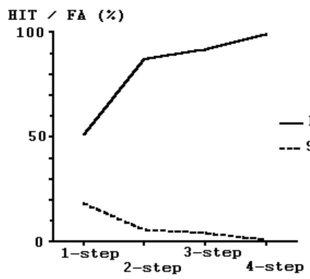

Figure 4. Effects of step-size on detection performance: Hit rate (solid lines) and false alarm rate (dashed lines) for each of the four step-size conditions, collapsed across all stimulus types (VOT10 - VOT50).

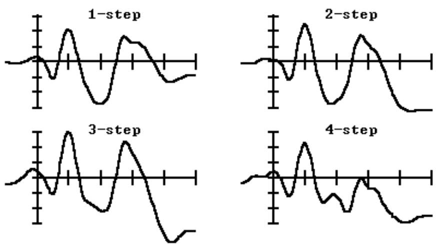

Figure 5. Effects of step-size on AEP waveforms: Grand mean average waveforms at Cz, plotted separately for each of four step-size conditions, collapsed across all stimulus types (VOT10 - VOT50).

For serial oder there was no effect for hit or or false alarm rate (see Fig. 6). N1-amplitude tended to be larger if the eliciting stimuli occurred toghether with stimuli of shorter VOTs (Fig. 7), but this effect only approached statistical significance (F(1,8) = 4.3, p = .072). There was, however, a significant interaction of TYPE x ORDER (F(1,8) = 6.99, p = .030) in the P2 latency range (see Fig. 7): If the eliciting stimuli occured together with stimuli of longer VOTs, a substantial reduction of the P2 amplitude elicited by deviant items relative to the P2 amplitude elicited by standard items could be observed. On the other hand, if the eliciting stimuli occurred together with stimuli of shorter VOTs, no such P2 effect was obtained. This finding was confirmed by additional ANOVAS, conducted separately for P2-amplitudes elicited by stimuli occuring together with stimuli of longer VOTs, and for P2-amplitudes elicited by stimuli occurring together with stimuli of shorter VOTs. In the former case, the P2-amplitudes elicited by standard and deviant items differed significantly (F(1,8) = 12.52, p = .008), while in the latter case, no such difference occurred (F < .05). Finally, serial order marginally influenced the N2b component, as evidenced by an almost significant three-way interaction of TYPE x ORDER x ELECTRODE (F(2,16) = 3.64, p = .062).

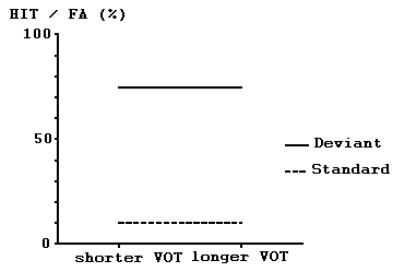

Figure 6. Effect of serial order on detection performance: Hit rate (solid line) and false alarm rate (dashed line) for the order conditions, collapsed across all stimulus types (VOT10 - VOT50).

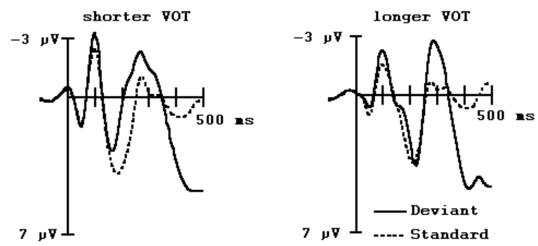

Figure 7. Effects of serial order on AEP waveforms: Grand mean average waveforms at Cz, plotted separately for deviant items (solid lines) and standard items (dashed lines) collapsed across all stimulus types (VOT10 - VOT50). Left panel: VOT of the stimulus eliciting the depicted waveform is shorter than the VOT of ist respective context stimulus (i.e. deviant items occurring in a context of standards with longer VOT, and standard items occurring in a context of deviants with longer VOT). Right panel: VOT of the stimulus eliciting the depicted waveform is longer than the VOT of its respective context stimulus.

4. DISCUSSION

The main goal of the present experiment was to test whether AEPs would provide a reliable measure of serial order effects in speech processing. To this purpose, a 5-step VOT continuum ([ba] to [pa]) was presented in form of an oddball design, i.e., in each block one of the 5 stimuli was presented frequently, and a second one was presented infrequently. All 20 pairings of the 5 stimuli were tested. The main effect of stimulus deviance on the N1- and N2b-component (i.e. enhanced negativities elicited by deviant items compared with standard items) indicates that the experimental manipulation was successful. Behavioral results indicate a strong influence of step-size (hit rate near 100% for the largest step-size, and only about 50% for the smallest). This effect probably also accounts for the finding of an effect of VOT on behavioral performance: Since with the central stimulus (VOT30) only step-sizes of 1 and 2 could be realized in the present study, the reduced detection performance for this stimulus probably has to be regarded as being due to the more difficult discrimination rather than to some property of the stimulus itself. However, this reasoning does not hold for the effect of VOT on the N1 amplitude: The fact that the N1 was maximal for the shortest VOT and minimal for the longest VOT cannot be explained in terms of a step-size artifact (as step-sizes were the same for both edge stimuli), and thus rather has to be regarded as a genuine VOT effect. The most likely explanation for this finding is that the initial burst and the following voice onset each generate a separate N1 component. With shorter VOTs, the temporal overlap of these components results in an enhanced total negativity, while with longer VOTs, the temporal gap between them results in a reduced total negativity.

The most interesting result of the present experiment is the finding that serial order of stimulus presentation had a strong influence on AEPs. If deviant stimuli were presented in a context of standard stimuli with longer VOTs, they exhibited a substantial reduction of the P2 component compared with the P2 elicited by the corresponding standard stimuli. If the same stimuli were presented in a context of standard stimuli with shorter VOTs, no such effect was observed. Again, this effect cannot be explained as an artifact of step-size or VOT, as neither step-size nor VOT had an influence on the P2 component. Visual inspection of the AEP waveforms suggests that the P2 reduction stems from an earlier onset of the N2b component for deviant stimuli with shorter VOTs than their context. One may speculate that if the context consisted of stimuli with long VOTs, then the occurrance of a deviant with a shorter VOT results in an early, 'active' disruption of the expectancy template. If, on the other hand, the context consists of stimuli with short VOTs, then the occurrance of a stimulus with longer VOT results in a later and more 'passive' detection of stimulus deviance. Consequently, one could assume that in the former case attention is drawn automatically to the deviant stimulus while it is still being processed, whereas in the latter case the deviant stimulus 'comes to attention' after it has been processed. This reasoning is in line with the finding that serial order had no effect on hit rate: Since the need to respond quickly was not specifically emphasized, it seems likely that participants responded only after the processing of each stimulus was complete. Although one could speculate that participants relied on different sources of evidence in order to make their decision (early active template disruption vs. late post-hoc deviance recognition), this hypothesis can not be proofed in the context of the present study. It seems promising, however, to investigate this issue more directly in a study aimed at investigating serial order effects under more restricted speeded response conditions.

ACKNOWLEDGMENT

This work was supported by the German Research Council (DFG grant no. Pi 200/2-2). The experiment was conducted at the Institute of Psychology, Munich University.

REFERENCES

[1] Näätänen, R. and Picton, T.W. 1986. N2 in automatic vs. controlled processes. In: McCallum, W.C., Zappoli, R. and Denoth, F. (eds.), *Cerebral psychophysiologies: Studies in event-related potentials*. Amsterdam, 169-186.

[2] Squires, N.K., Squires, K.C. and Hillyard, S.A. 1975. Two varieties of long-latency positive waves evoked by unpredictable stimuli in man. *Electroencephalography and Clinical Neurophysiology*, 38, 387-401.

[3] Greenberg, H.J. and Graham, J.T. 1970. Electroencephalographic changes during learning of speech and nonspeech stimuli. *Journal of Verbal Learning and Verbal Behavior*, 9, 274-281.

[4] Dorman, M.F 1974. Auditory evoked potential correlates of speech sound discrimination. *Perception and Psychophysics*, 15, 215-220.

[5] Aaltonen, O., Niemi, P., Nyrke, T. and Tuhkanen, M. 1987. Event-related brain potentials and the perception of a phonetic continuum. *Biological Psychology*, 24, 197-207.

[6] Sams, M., Aulanko, R., Aaltonen, O. and Näätänen, R. 1990. Event-related potentials to infrequent changes in synthesized phonetic stimuli. *Journal of Cognitive Neuroscience*, 2, 344-357.

[7] Sharma, A., Kraus, N., McGee, T., Carrel, T. and Nicol, T. 1993. Acoustic versus phonetic representation of speech as reflected by the mismatch negativity event-related potential. *Electroencephalography and Clinical Neurophysiology*, 88, 64-71.

[8] Phillips, C., Marantz, A., McGinnis, M., Petetsky, D., Wexler, K., Yellin, E., Poeppel, D., Roberts, T. and Rowley, H. 1995. Brain mechanisms of speech perception: A preliminary report. *MIT Working Papers in Linguistics*, 1-40.

[9] Fechner, G.T. 1860. *Elemente der Psychophysik*. Leipzig.

[10] Stott, L.H. 1935. Time order errors in the discrimination of short tonal durations. *Journal of Experimental Psychology*, 18, 741-766.

[11] Piroth, H.G. and Tillmann, H.G. 1987. An order effect in pulse train discrimination as a case of time order error. *Proceedings 11th ICPhS*, Tallinn, vol. 5, 50-53.

[12] Uselding, D.K. 1977. A temporal order effect in voice onset time discrimination. *Language and Speech*, 20, 366-376.

[13] Schiefer, L. and Batliner, T. 1991. A ramble round the order effect. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, 29, 125-180.

[14] Klatt, D.H. 1980. Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, 67, 971-995.