

# TRANSITIONAL PROBABILITY AND PHONEME MONITORING

*James M. McQueen*

Max-Planck-Institute for Psycholinguistics  
Nijmegen, The Netherlands

*Mark A. Pitt*

Ohio State University  
Columbus, Ohio, U.S.A.

## ABSTRACT

Two phoneme monitoring experiments examined the influence of Transitional Probability (TP) on phoneme recognition. Target phonemes appeared at the end of Consonant-Vowel-Consonant (CVC) syllables, or as the first element of coda clusters in CVCC syllables. Reliable TP effects were found only for targets in CVCC syllables. The TPs both into and out of the targets influenced listeners' ability to detect them in CVCCs. Furthermore, targets were more difficult to detect in CVCCs than in CVCs. TP may only influence segment recognition when that segment is more difficult to recognise, as when it occurs in a cluster.

## 1. INTRODUCTION

The relative likelihoods of phonemes appearing after other phonemes (i.e., their frequency of co-occurrence or Transitional Probability [TP]) may be used by listeners in speech recognition. In the absence of clear information in the signal, TP could be used to predict which phonemes are more likely in a given context. TP could also be used to assist in the process of segmentation of continuous speech, since, for example, sequences of phonemes which do not occur within words cue the location of word boundaries. Evidence is steadily accumulating that TP plays a role in segmentation. For example, nine-month-old infants are sensitive to the sequential patterns of speech sounds in their native language [1], suggesting that they may use TP information when first segmenting speech in order to learn words. It has also been shown that consonant clusters with TPs of zero (phonotactically-illegal clusters) provide a segmentation cue for adults; Dutch listeners find it easier to detect words embedded in nonsense strings when the embedded words are aligned with syllable boundaries determined by phonotactics (such as *vel*, *skin*, in *vel.brul*) than when they are misaligned (such as *vel* in *velm.rul*, [2]). Adults learning an artificial language also show sensitivity to the relative likelihoods of different phoneme sequences in that language [3].

These results suggest that listeners can use TP information in segmentation. Some evidence also suggests that TP can contribute to the recognition of speech sounds. For example, phonemes can be detected more easily when they occur in high frequency syllables than when they occur in low frequency syllables [4]: high frequency syllables will tend to have high TPs; low frequency syllables will tend to have low TPs. It has also been shown that TP influences the categorisation of ambiguous phonemes [5]. But there has been no systematic analysis that has investigated the influence of TP on unambiguous phoneme perception.

The present study therefore sought to examine in detail TP effects on phoneme recognition. The influence of TPs both into and out of target phonemes in monosyllabic Dutch nonwords was examined in two phoneme monitoring experiments. Two types of monosyllables were examined. Monitoring for consonantal targets in the coda of CVC items was examined through the factorial manipulation of TP in both the transition into the vowel (the CV transition) and the transition out of the vowel (the VC transition). Monitoring for targets that formed the first element of coda clusters in CVCC items was also tested, allowing for a comparison between target detection in clusters and that in singleton codas. The use of CVCC items also allowed for the influence of TPs into the target phoneme (CVC transitions) to be compared with the influence of TPs out of the target phoneme (CC transitions). In the CVCC items, CVC and CC TPs were therefore manipulated factorially.

## 2. EXPERIMENT 1

### 2.1. Method

**Materials.** Two sets of nonword materials were constructed, one consisting of CVC items, and one of CVCC items. Within the CVCs, there were 80 experimental items with target consonants in their codas, 20 in each of four TP conditions, as shown in Table 1.

The probabilities were computed using the CELEX database. TPs were computed for every diphone combination in the Dutch language, based on a count (weighted by word frequency in a corpus of 42 million words) over the phonological transcriptions of the 124,136 headwords in the database. All CVC combinations which did not form Dutch words were then ranked according to the TPs of their CV and VC portions. Items were selected for the high TP conditions if TPs fell in the highest quartile of the range, and for the low TP conditions if TPs fell in the lowest quartile.

Lists of CVC nonwords were then constructed for each target-bearing item. They were placed in the penultimate position in lists with either three or four other CVC nonwords, none of which contained the target consonant. In addition to these 80 lists, 20 lists of CVC nonwords were constructed which contained no instances of the specified target phoneme. A further 24 lists of CVC nonwords were made in which the specified target phoneme did occur, as the coda in a filler nonword, but these nonwords appeared in non-penultimate positions in the lists (thus making the position of target-bearing items in lists unpredictable). The filler nonwords used in all these lists were selected from the interquartile ranges of the CV and VC TPs (i.e., no particularly high or low TPs were used in the fillers). Eight practice lists were also made.

CVC Items		CV Transition	
		high	low
VC Transition	high	weel	soel
	low	houl	feul
CVCC Items		CVC Transition	
		high	low
CC Transition	high	weels	feuls
	low	weelm	feulm

**Table1:** Design. Examples are given for CVC and CVCC items, each in four conditions, defined by high or low probability of their component transitions. The target phoneme in all examples was [l].

A similar procedure was used to construct the CVCC items. There were again 80 experimental items, all with target consonants as the first element in the coda cluster. Twenty items were selected in each of four TP conditions, as shown in Table 1 (CVC transitions were selected such that both CV and VC transitions had either high or low probability). Transitions were based on the same CELEX count, with the same selection procedure. These items were presented in penultimate positions in lists of either four or five CVCC nonwords. There were also 20 lists with no target, 24 lists with targets in the first half of the coda of CVCC items which did not appear in penultimate list position, and eight practice lists.

**Subjects and Procedure.** Fifty-five student volunteers were paid to take part in the experiment. They were split into two groups: 27 heard the CVC lists and 28 heard the CVCC lists. All stimulus lists were recorded onto DAT tape by a female native speaker of Dutch. The lists were then transferred to computer and items were measured using a speech editor. Subjects were tested in individual sound-proof booths. They were instructed to listen to lists of nonsense words, and to press a button as fast as possible if any nonword contained a pre-specified target sound. The targets were presented visually, in upper-case letters on a computer screen in front of the subject (NG was used to represent the velar nasal; otherwise single letters corresponding to each of the other targets were used: M, N, L, R, P, T, K, F, S and G [for the velar fricative]). Subjects were informed that a new target would appear for 1.5 seconds on the screen before each list was heard. Presentation of materials was controlled by computer, with the auditory lists being played directly from hard disk. The lists were presented binaurally, over headphones. The computer also timed and stored Reaction Times (RTs). Prior to statistical analysis, RTs were adjusted so as to measure from the onset of each target.

## 2.2. Results and Discussion

Mean RTs and mean error rates are given in Table 2. Analyses of Variance (ANOVAs) were performed on both the latency and the

error data, treating both subjects (F1) and items (F2) as the repeated measure. Overall analyses showed that targets in CVC items were detected more rapidly than those in CVCC items (by 79 msec, on average:  $F1(1,53) = 15.24$ ,  $p < 0.001$ ;  $F2(1,152) = 42.36$ ,  $p < 0.001$ ) and that targets in CVCs were detected more accurately than those in CVCCs (by 8%, on average:  $F1(1,53) = 19.02$ ,  $p < 0.001$ ;  $F2(1,152) = 16.30$ ,  $p < 0.001$ ). Subsequent analyses were performed on CVC and CVCC items separately. In the CVC analyses, there were no reliable effects, neither by subjects nor items, in the RT data. In the error data, there was a main effect of VC transition: overall, subjects were 3% more errorful on CVC items with low VC transitions. But this effect was only significant by subjects ( $F1(1,26) = 8.16$ ,  $p < 0.01$ ;  $F2(1,76) = 3.66$ ,  $p = 0.06$ ).

In the CVCC analyses, however, there were reliable effects of TP. In RT, subjects were faster to detect targets when the CC transition out of the target phoneme was highly probable (457 msec, on average) than when the CC transition had low probability (486 msec, on average). This effect was only reliable by subjects:  $F1(1,27) = 15.72$ ,  $p < 0.001$ ;  $F2(1,76) = 2.47$ ,  $p > 0.1$ . There was also an interaction between the CVC and CC transition factors, with the disadvantage for low CC TP items due entirely to items with low CVC transitions ( $F1(1,27) = 11.78$ ,  $p < 0.005$ ;  $F2(1,76) = 3.90$ ,  $p = 0.05$ ). Pairwise t-tests showed that the CC effect within the low CVC items (57 msec, on average) was reliable:  $t1(1,27) = 4.64$ ,  $p < 0.001$ ;  $t2(1,19) = 2.38$ ,  $p < 0.05$ . In errors, subjects were less accurate in items with low CVC TPs (24% misses, on average) than in items with high CVC TPs (8%, on average):  $F1(1,27) = 91.66$ ,  $p < 0.001$ ;  $F2(1,76) = 22.88$ ,  $p < 0.001$ . There was again an interaction between the CVC and CC TP conditions, though this was only significant by subjects. The disadvantage for low CVC TP items was more marked in those items which had low CC TPs (18%, on average) than in those which had high CC TPs (15%, on average):  $F1(1,27) = 14.06$ ,  $p < 0.001$ ;  $F2 < 1$ . Pairwise t-tests showed that performance was reliably poorer on items with low CVC TPs than on items with high CVC TPs, both when the CC transition was highly probable ( $t1(27) = 5.48$ ,  $p < 0.001$ ;  $t2(38) = 2.72$ ,  $p < 0.05$ ) and when the CC transition was of low probability ( $t1(27) = 8.70$ ,  $p < 0.001$ ;  $t2(38) = 4.30$ ,  $p < 0.001$ ).

CVC Items		CV Transition	
		high	low
VC Transition	high	387 (7%)	402 (7%)
	low	392 (11%)	386 (9%)
CVCC Items		CVC Transition	
		high	low
CC Transition	high	472 (11%)	442 (26%)
	low	472 (5%)	499 (23%)

**Table 2:** Experiment 1. Mean RT (in milliseconds) and mean miss rates (in parentheses), for the CVC and CVCC items.

Experiment 1 therefore failed to show any reliable TP effects on CVC items. Two effects emerged in the CVCC items, however. Targets with high CVC TPs were detected more accurately than those with low CVC TPs - targets were detected more accurately when the preceding consonant and vowel made them more probable continuations; and, within items with low CVC TPs, there was an effect of the CC transition - targets were detected more rapidly when the consonants following them were more likely.

Why might TP effects be limited to CVCC items? One possibility is that TP effects tend only to emerge when target consonants are more difficult to detect (CVCC performance was worse than CVC performance). If this were the case, TP effects could perhaps emerge in CVC items if the task were made more difficult. One way to do this is to vary the location of the target consonant in the stimuli. Target location was fixed in Experiment 1. Subjects could learn quickly that targets would always appear immediately after the vowel. Furthermore, subjects heard only CVC items or only CVCC items. These factors could allow subjects to focus attention on target locations and ignore contextual information. This focussing of attention could weaken any effects of the surrounding context, such as might be due to TP.

Experiment 2 therefore tested for TP effects under conditions where target position was unpredictable. Listeners were forced to process each item more fully: the CVC and CVCC items from Experiment 1 were mixed together, such that subjects would hear targets in both singleton codas and in coda clusters. The location of the target phonemes was further varied in additional filler items, such that targets also appeared in onsets (in both CVC and CVCC items), and as the final consonant in the codas of CVCC items.

### 3. EXPERIMENT 2

#### 3.1. Method

**Materials.** The 160 target-bearing experimental items from Experiment 1 (80 CVC and 80 CVCC) were used. These items were split into two mixed sets, each comprising 40 CVC and 40 CVCC items. They were then placed in the penultimate position in lists of other CVC and CVCC nonwords, selected at random from the fillers used in Experiment 1, with the constraint that each list contained at least one nonword of each CV-structure.

Further lists, also consisting of a mixture of CVC and CVCC filler items taken from Experiment 1, were constructed, some containing a target phoneme, and some not. Those with targets were split both according to the list position of the target-bearing item (making position of target-bearing items unpredictable), and according to the target's position in the item (so that that too was unpredictable). Two sets of lists were made, each consisting of: 40 experimental lists with targets in the coda of CVC items; 40 experimental lists with targets as the first element in the coda cluster of CVCC items; 80 filler lists with targets in the onset of items (half CVC and half CVCC); 20 filler lists with targets as the final consonant in CVCC items; 30 lists containing no instances of the target consonant; and eight practice lists.

**Subjects and Procedure.** Forty-eight student volunteers were paid to take part. The new lists were re-recorded, by the same speaker. The preparation and running of the experiment were the same as in Experiment 1. Two subjects were excluded from the analysis because of high error rates (over one third of all targets missed).

#### 3.2. Results and Discussion

Mean RTs (from target onset) and mean error rates are given in Table 3. Overall ANOVAs again showed that targets in CVCCs were detected more slowly than those in CVCs (by 59 msec, on average:  $F(1,44) = 72.34, p < 0.001$ ;  $F(1,152) = 28.06, p < 0.001$ ). The difference in error rates (CVC performance 2% more accurate, on average than CVCC performance) was not significant. Subsequent analyses were then performed separately for the CVC and CVCC items. In the CVC items, a very similar pattern of results was obtained to that found in Experiment 1. There were no reliable effects, neither by subjects nor items, in the RT data. In the error data, there was a main effect of VC transition: overall, subjects were 4% more errorful on CVC items with low VC transitions. But this effect was again only significant by subjects ( $F(1,44) = 8.29, p < 0.01$ ;  $F(2, 1) < 1$ ). It appears that the failure to find reliable effects of TP in CVC items was not a consequence of the fixed target position in Experiment 1. Exactly the same weak tendency was found here.

In the CVCC items, as in Experiment 1, TP effects were observed. In RT, subjects showed a tendency to respond more quickly to targets in items with low CVC TPs (439 msec, on average) than to targets in items with high CVC TPs (469 msec, on average). This inverse TP effect, however, was only significant by subjects:  $F(1,44) = 13.52, p < 0.001$ ;  $F(2,1,76) = 1.13, p > 0.2$ . A TP effect in the predicted direction was observed in the error analysis. Subjects missed twice as many targets in items with low CVC and CC transitions than in items in any of the other three conditions. This pattern was reflected in main effects of both CVC transitions ( $F(1,44) = 14.62, p < 0.001$ ;  $F(2,1,76) = 3.05, p = 0.08$ ) and CC transitions ( $F(1,44) = 15.12, p < 0.001$ ;  $F(2,1,76) = 3.28, p = 0.07$ ),

CVC Items		CV Transition	
		high	low
VC Transition	high	395 (13%)	389 (11%)
	low	395 (16%)	402 (16%)
CVCC Items		CVC Transition	
		high	low
CC Transition	high	468 (12%)	445 (13%)
	low	471 (13%)	433 (24%)

**Table 3:** Experiment 2. Mean RT (in milliseconds) and mean miss rates (in parentheses), for the CVC and CVCC items.

and in an interaction of these two effects ( $F(1,44) = 10.49$ ,  $p < 0.005$ ;  $F(1,76) = 2.06$ ,  $p > 0.1$ ).

Pairwise t-tests painted a similar picture. The only significant differences involved the low CVC and low CC TPs. They were responded to less accurately than those in items with low CVC TPs but high CC TPs (an 11% difference, on average;  $t(45) = 4.46$ ,  $p < 0.001$ ,  $t(38) = 1.93$ ,  $p = 0.06$ ), those with low CC TPs but high CVC TPs (also an 11% difference, on average;  $t(45) = 4.19$ ,  $p < 0.001$ ,  $t(38) = 1.95$ ,  $p = 0.06$ ), and those with high CVC and CC TPs (a 12% difference, on average;  $t(45) = 4.90$ ,  $p < 0.001$ ;  $t(38) = 2.29$ ,  $p < 0.05$ ). Note that although responses to the low CVC/low CC TP condition were the most errorful, they were also the fastest. But the RTs on these items were not reliably faster (in t-tests) than on items in any of the other three conditions. The high error rate on the low CVC/low CC TP condition was not simply the consequence of a speed-accuracy trade-off.

Performance on the CVCC items in Experiment 2 therefore showed a TP effect. The effect was not the same as in Experiment 1, where an effect of the CVC transition was seen in error rates, and a conjoint effect of both the CVC and CC transitions was seen in RT. Here, a conjoint effect of both CVC and CC TPs (transitions both into and out of the target phoneme) was observed in error rates. Targets with the lowest TPs were detected reliably less accurately than those with the highest TPs.

## 4. GENERAL DISCUSSION

Listeners failed to show any reliable sensitivity to the probability of transitions between phonemes in their ability to detect final consonants in CVC items. Sensitivity to TP, however, was observed when listeners had to detect phonemes in the first part of the coda cluster in CVCC items. The likelihood of transitions both into and out of the target phoneme influenced performance.

Although TP effects were observed in CVCCs, they were not always fully reliable, and they differed across the two experiments. These factors suggest, along with the failure to find any effects in CVCs, that TP has a limited role to play in phoneme recognition, at least for speech sounds recorded under high-quality conditions. This is consistent with the view that human speech recognition is highly data-driven. Given good information in the input, the recognition system will tend to use it. Under more difficult listening conditions, the recognition system may depend more heavily on contextual factors, such as the relative likelihoods of different segments in different contexts. If the materials from the current experiments were degraded (by presentation in noise or by low-pass filtering), stronger effects of TP could perhaps be found.

Data from the phonetic categorisation task support this argument. Robust TP effects can be seen in the categorisation of ambiguous phonemes. For example, the labelling of ambiguous phonemes [ʔ], ranging between [l] and [r], is influenced by the phonotactic legality of the cluster containing those phonemes [5]. In [tʔi], [r] responses are more likely, and in [sʔi], [l] responses are more likely. Phonotactically-illegal sequences have extreme TPs: [l] has

effectively zero probability between [t] and [i]. Such effects indicate that the recognition system, when confronted with extreme perceptual uncertainty (i.e., ambiguous phonemes), can make use of TP information to assist in the identification of speech sounds.

The present results also support this argument. The TP effects occurred when targets were in clusters, in the CVCC items. The quality of information specifying a consonant is likely to be poorer when that consonant appears in a cluster than when it appears as a singleton coda. Targets in CVCCs were in fact detected more slowly than those in CVCs in both experiments, and they were detected less accurately than those in CVCs in Experiment 1. It has previously been shown in syllable-initial phoneme monitoring that targets can be more difficult to detect in clusters than as singletons [6]. A phoneme's likelihood of occurrence may exert a stronger influence on recognition when there is poorer information in the signal specifying that phoneme, as when the phoneme is either perceptually ambiguous or in a consonant cluster. It appears that TP can be used both in the segmentation of continuous speech and in the recognition of speech sounds.

## 5. ACKNOWLEDGEMENT

We would like to thank Arie van der Lugt and Dirk Janssen for their help in the preparation and running of these experiments. Partial funding for this research was provided by the International Human Frontier Science Program. We gratefully acknowledge their support. Correspondence can be addressed to J.M. McQueen, MPI for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands, or by email to either author (James.McQueen@mpi.nl or Pitt.2@osu.edu).

## 6. REFERENCES

1. Jusczyk, P.W., Luce, P.A., and Charles-Luce, J., "Infants' sensitivity to phonotactic patterns in the native language", *J. Memory and Language*, Vol. 33, 1994, pp. 630-645.
2. McQueen, J.M., and Cox, E., "The use of phonotactic constraints in the segmentation of Dutch", *Proc. Eurospeech '95*, Vol. 3, 1995, pp. 1707-1710.
3. Saffran, J.R., Newport, E.L., and Aslin, R.N., "Word segmentation: the role of distributional cues", *J. Memory and Language*, in press.
4. Pitt, M.A., and Samuel, A.G., "Lexical and sublexical feedback in auditory word recognition", *Cognitive Psych.*, Vol. 29, 1995, pp. 149-188.
5. Massaro, D.W., and Cohen, M.M., "Phonological constraints in speech perception", *Perception & Psychophysics*, Vol. 34, 1983, pp. 338-348.
6. Cutler, A., Butterfield, S., and Williams, J.N., "The perceptual integrity of syllabic onsets", *J. Memory and Language*, Vol. 26, 1987, pp. 406-418.