LIMITED LOOKAHEAD IN SPEECH PRODUCTION

Sieb G. Nooteboom
Trans 10
Research Institute for Language and Speech (OTS)
3512 JK UTRECHT
The Netherlands

Many years ago, at a meeting of the Acoustical Society of America, Katherine Harris asked me whether declination of pitch in speech could provide information on the amount of preprogramming in speech production. I do not recall my answer, but it must have been very unsatisfactory. Below I make a second attempt, considering also some other sources of empirical evidence on preprogramming in speech.

1. Introduction

Wolfgang Amadeus Mozart, reporting about a new composition taking form in his mind, said: "The work grows; I keep expanding it, conceiving it more and more clearly until I have the entire composition finished in my head though it may be long. Then my mind seizes it as a glance of my eye a beautiful picture or a handsome youth. It does not come to me successively, with various parts worked out in detail, as they will later on, but in its entirety that my imagination lets me hear it" (Hadamard, 1945, cf. Penrose, 1990).

Obviously, composing takes time, but "hearing" the end result as "inner music" for Mozart was immediate, all the music being mentally present simultaneously to be written up or performed. Analogously, we can imagine a gifted orator who "composes" a long speech in his mind, where the whole speech sits waiting, as a massive phonetic plan, to be either further refined by mental editing or to be spoken. People like Mozart or our imaginary orator (not fully imaginary, however, I have known at least one man who seemed to prepare his speeches as Mozart did his compositions), are rare however. Most composers and most orators have either to use script as an external memory during preparation or to improvise while performing. One can, of course, also learn a whole composition or long speech by heart before performing, but this rarely seems to lead to the experience described by Mozart. Learning by heart involves all kinds of memory tricks that help in retrieving all details in the right order from long term memory. It does not normally lead to a whole composition or speech being simultaneously present to the conscious mind.

Yet, during normal spontaneous speech production, there must be something in the mind before it can be produced. Interestingly, it does not help very much to look into your own mind while speaking, and ask "What 's there? What is my phonetic plan?
How many words does it contain? Which are the details of speech taken care of by my phonetic plan, and which are the ones only added by the output mechanisms?". During actual speech production the mind does not seem to have the ability to keep track of its own workings to such an extent that the above questions can be answered. The same is true for speech that is imagined but not spoken. Although most people I asked about it, know the

experience of inner speech, no one could give me any detailed answers as to, for example, the amount of inner speech that is simultaneously present to the conscious mind.

This brings us back to observable overt speech, hopefully providing a window on the mental processes underlying speech production. In doing so we use our mental capacities to study our own mental processes as if they belonged to some alien species. Let us look at the following quotation from Chapter 1 of the Speech Science Primer by Borden and Harris (1980):

"It seems probable that chunks of the message are briefly stored in a buffer (temporary storage) ready for output. The chunks are perhaps of sentence length or phrase length. Evidence for this storage comes from slips of the tongue. The fact that people make mistakes such as 'He cut the knife with the salami', Victoria Fromkin's example, indicates the existence of such a buffer in order for the speaker to have substituted what should have been the last word for the fourth-to-last-word word". The quotation is revealing. "It seems probable.." adequately suggests that it is not easy to say anything definite about the contents of the temporary storage supposedly preceding overt speech. The suggestion that the "chunks" might be of sentence length or phrase length, although perhaps somewhat naive in not taking into account the possibility of incremental processing during phrase or sentence production, precisely in its naivety is suggestive of how slippery this area of research is. The quotation also shows how phenomena in overt speech might be used as hints about the underlying mental processes. The error of speech quoted from Fromkin would suggest that the phonetic plan contains at least four words, in this case "salami with the knife", unintendedly realized as "knife with the salami". But note that we cannot be sure that at the level of programming where the error occurs the function words "with the" were already selected. It is at least imaginable that there exists a level of mental programming at which content words are selected together with all the semantic, syntactic, and morphological information necessary for later insertion of function words and affixes. If so, the error would be suggestive of only two words being simultaneously present in the mental programme at that level, or of one word "lookahead" as it is defined by Levelt (1989). Function words and affixes would then be comparable to the "various parts worked out in detail" that do not enter into Mozarts timeless mental picture of a symphony.

Errors of speech are not the only source of hints about the size of the mental programme for speaking. Acoustic/phonetic details such as vowel duration or the course of pitch, depending on aspects of the message yet to come, may also tell us something about the size of lookahead during speaking. Below I will discuss some relevant observations from four different areas: starting frequency and declination of pitch, accent lending rises and falls in Dutch intonation, anticipatory shortening, and errors of speech. At some points I will take issue with statements or arguments in Chapter 10 of Levelt's admirable book "Speaking. From intention to articulation" (1989). Throughout this book Levelt assumes, as many others have done, that the mental production of speech is a multi-stage process, including for example separate stages for conceptualizing, grammatical encoding, phonological encoding, and articulating, and that it is incremental, meaning that the next stage of processing does not have to wait until the output of the preceding stage has been completed. Rather, processing at each successive stage starts on the basis of partial, incomplete output from the preceding stage. Therefore processing at different stages overlaps in time. Much lookahead is necessary lookahead if we assume incremental rather than if we assume serial processing. In fact, Levelt throughout his book assumes that the minimally necessary lookahead at the level of the "phonetic plan", being the output of phonological encoding and the input of articulation, is only one word, meaning that the phonetic plan does not have to contain more than two words at a time, the word under production and the next word. Occasionally more lookahead may occur, for example during reading aloud, or perhaps in very gifted speakers with an unusual mental span of attention, but this is not necessary for the production of fluent speech. More lookahead may, according to Levelt, make speech more esthetically pleasing, though. The reader may notice that it would not be easy to show that Levelt's position is incorrect. There is very little systematic acoustic/phonetic evidence relating to spontaneous speech. Any evidence of more than one word lookahead can be explained as being occasional, and not indicative of what is minimally necessary for the production of fluent speech. What I set out to do below, then, is not to argue that on the level of the phonetic plan more lookahead is minimally necessary for the production of acceptable spontaneous speech, but to review some empirical evidence as to the amount of lookahead actually present during normal speech production. It will also become apparent that at some points, notably with respect to spontaneous speech, empirical evidence is insufficient. I will assume that on the level of the phonetic plan, resulting from phonological encoding, all words are spelled out from early to late as sequences of segments, together with temporal structures and pitch patterns. The output of grammatical encoding, containing a sequence of lexical items with grammatical structure, is not necessarily fully specified as a sequence of all morphemes to be produced.

2. Declination

During the course of coherent stretches of speech, often called intonational phrases, the pitch gradually drifts down. This phenomenon is known as declination (Cohen and 't Hart, 1967). It has also been observed that declination gets less steep with increasing length of the intonational phrases concerned (Ohala, 1978; Cooper and Sorensen, 1981; De Pijper, 1983; 't Hart, Collier and Cohen, 1990). As mentioned by Levelt (1989, p. 400), this has been taken as evidence for preprogramming over the length of an intonational phrase. Levelt points out, however, that, firstly, declination is much more clearly present in reading than in spontaneous speech, and secondly, that the causal relation may be inverse: "If a speaker, for whatever reason, makes his pitch decline rapidly, he will sooner feel the urge to rest. This may induce him to take an early break option. Consequently, the running intonational phrase will be a short one". There is something Levelt does not mention, however. Both in reading aloud and in spontaneous speech pitch tends to move towards the same speaker specific value at the end of intonational phrases, but to show different values at the beginning of intonational phrases, where pitch is higher as the intonational phrases to be produced are longer (Cooper and Sorensen, 1981; 't Hart, Collier and Cohen, 1990). This suggests that speakers adapt their starting frequency to the length of the intonational phrase to be spoken, or, conversely, that speakers adapt the length of the intonational phrase to both the starting frequency and the slope of declination.

Willems (1983) measured declination slopes in 35 read-out British-English sentences and in 35 spontaneous British-English utterances varying in duration from 0.6 to 6.3 seconds. He compared measured slopes with slopes predicted from the following formulas:

For $t < 5$: $D = -11 / t + 1.5$

For $t > 5$: $D = -8.5 / t$

in which D is the slope to be calculated and t is duration in seconds. The formulas predict a slope of - 5 semitones/second for the shortest utterance and of - 1.35 semitones/second for the longest utterance. End frequencies were set at speaker specific fixed values. Average differences between predicted and measured slopes were - 0.51 semitones/second for reading aloud and - 0.3 semitones/second for spontaneous utterances. Standard deviations were 0.79 semitones/second for reading aloud and 2.05 semitones/second for spontaneous speech.

't Hart, Collier, and Cohen (1990) conclude from Willems' data "...that speakers apply a certain amount of preplanning: if the duration of the utterance they are about to produce is known in advance, they can choose a start frequency and a slope suitable to finish at their individual end frequency. Estimating the duration can, understandably, be done fairly accurately if the speaker is reading from text. The fact that such a preprogramming is less successful in individual cases of spontaneous speech does not entirely rule out its occurrence". I would add that, given that on the average the effect shows up in a corpus of only 35 spontaneous utterances, there is at least the suggestion that such preplanning is not exceptional. Unfortunately, the corpus is too limited and the variance for spontaneous utterances too big to say anything definite on the size of the supposed lookahead. In principle, this could be remedied however, in future simular studies using more extensive data bases. Let us assume for the moment that the correlation between starting frequency and length of intonational phrases for individual speakers would be significant upto a length of intonational phrases of three or four words.

What then about Levelt's inverse causal relation? Logically we could argue that, if a speaker for whatever reason starts high, he would then be inclined by whatever mechanism not to let his pitch drift down very rapidly, and thus he would not be forced by his pitch getting too low to use an early optional breakpoint, and often go for a later optional breakpoint. This would explain the correlation between starting frequency, slope of declination and sentence length without the assumption of considerable lookahead. There is, however, a problem with this line of argumentation. Many intonational phrases are also sentences, or at least complete utterances, the shortest being only one word in length. If we believe in Levelt's inverse causal relation, we have also to assume that, in case a speaker for whatever reason starts on a relatively high pitch with a relatively slow declination, this causes him to make a longer sentence. So now we have to believe that the number of words in an utterance or spoken sentence is determined by the relative height of pitch and the relative slope of declination at the moment speaking starts, and not by the conceptual intentions of the speaker. I would not be surprised if even Levelt would find this hard to believe. Levelt's inverse causal relation is hard to disprove, but rather implausible.

Summarizing: both starting frequency and slope of declination in intonational phrases correlate with the length of the phrase. Assuming, according to Levelt's line of argumentation, that phrasal length is caused by high starting frequency and slope of declination, forces us also to assume that the number of words in utterances consisting of one intonational phrase is caused by the way declination is programmed. This seems highly improbable. In principle, starting frequency and slope of declination can be used as indicators of the amount of lookahead at the level of the phonetic plan. For reading aloud, available evidence suggests that lookahead corresponds to a stretch of speech lasting a number of seconds, and containing more than a few words. There are to my knowledge no studies available from which the amount of actual lookahead can be estimated with any precision. For spontaneous speech, data on the correlation mentioned are too limited and variable to give any estimate of the lookahead involved. In principle, this could be remedied by further research on larger data bases. In such research one should focus on complete utterances, consisting of one intonational phrase.

3. The hat pattern

In Dutch one of the possible pitch configurations consists of an accent lending rise followed by an accent lending fall. This configuration is called the "hat pattern" (Cohen and 't Hart, 1967; 't Hart, Collier and Cohen, 1990). In normal emotionally and attitudinally neutral utterances, an accent lending rise has to be followed by an accent lending fall: what goes up has to come down. So the very moment a speaker makes an accent lending rise, he must, in order to produce a correct hat pattern, have sufficient lookahead to know that there will be an accented word within the same intonational phrase to be marked with an accent lending fall. The distances between such rises and falls in spontaneous speech may therefore be indicative of lookahead in the speech programme, as pointed out by Levelt (p. 405): "Lookahead is a condition for producing the hat pattern". He also points out that "the hat pattern is a more likely pitch contour when two accents are to be made in close succession", thus suggesting that in actual practice lookahead is fairly limited. Of course, what we need here are real data. Some relevant data can be found in Collier (1972) who among other things counted the distances between accent lending rises and falls in hat patterns in a corpus of 750 spontaneous Dutch utterances. Unfortunately for the present purpose, distances were calculated in syllables, not in words. Below distances are recalculated in words on the basis of the average word length of 1.8 syllables in Dutch spontaneous speech. Sixty-five percent of Collier's utterances contained a hat pattern. In 39 % of all hat patterns rise and fall fell compulsorily on the same syllable and thus on the same word because there were no later accented words in the utterance that could attract a fall. Of the remaining cases, where speakers had an option to produce the hat pattern over more than one word, 46 % still had rise and fall on the same word, with no necessary lookahead. The next and last accent in such cases was produced by a new rise-fall configuration. This may indicate that speakers tend to avoid the necessity of considerable lookahead. Of all cases where speakers did take the option of a hat pattern over more than one word, 32 % had rise and fall on two consecutive words, meaning that there was one word lookahead, 12 % had a rise and fall span of three words, indicating two words lookahead, and the remaining 10 % covered estimated spans of more than four words, suggesting more than three words lookahead. Thus in 22 % of these cases, corresponding to 8.5 % of the utterances in the entire corpus of 750 utterances, lookahead had to be at least more than one word.

These numbers can be taken to support Levelt's idea that generally lookahead is not necessarily more than one word, "generally" meaning here in c. 92 % of utterances. On the other hand one can argue that it is only possible to say anything about the lookahead involved in those cases where speakers had an option of producing a hat pattern over more words. And of those cases 22 % show a minimum lookahead of more than one word, suggesting that a lookahead of more words, although still not attested in the majority of cases, is far from exceptional. Of course, we cannot be sure that if speakers decline to take the option of an extended hat pattern, they do so because of lack of necessary lookahead. They may have other reasons, for example of a melodical nature. If that were generally the case, we could only say something about the lookahead in those cases where people are so kind to take the option of making an extended hat pattern. And of those cases 60 % show a minimum lookahead of one word, 22 % of two words, and 18 % of more than two words. So in all, 40 % of those cases where we really can say something about the minimum lookahead required, show a lookahead of more than a single word. But there seems to be no way in which we can make out whether this is representative of the amount of lookahead at all other moments during speech production.

Assuming that Collier's corpus is representative of Dutch spontaneous speech in general, we may conclude that in at least one out of every twelve Dutch spontaneous utterances lookahead is more than one word. For eleven out of twelve utterances there is no such evidence, one way or the other.

4. Anticipatory shortening

Anticipatory shortening is a well known and systematically occurring phenomenon in speech: the segments of a syllable, particularly a stressed syllable, become shorter as more syllables follow within the same word (Lindblom, 1968; Nooteboom, 1972). Anticipatory shortening is not limited to the level of words. The segments of a stressed word also shorten as the number of syllables or words coming later in the same intonational phrase increases (Nakatani, O'Connor and Aston, 1981, De Rooij, 1979). This suggests that the number of following words that still contribute to the shortening of the first syllable might be indicative of the amount of lookahead. Levelt, discussing this phenomenon, here again resorts to the inverse causal relation: "..one might conjecture that the speaker "blindly" increases the duration of successive stressed syllables till he reaches the end of the intonational phrase, and that he then resets the durational parameter to the initial value for the next phrase" (Levelt, 1989, p. 390). Levelt then continues: "Another possible explanation involves nuclear stress. Phrase-final stresses naturally "grow" toward the end of the sentence, owing to the mechanisms of the Nuclear-Stress Rule. And more heavily stressed syllables tend to be longer". In both explanations it is implied that durations of stressed syllables spoken in isolation are shortest, and that they get longer as more material is added in front of the syllable concerned: if a speaker, by lack of sufficient lookahead, does not know how many words are following, his realization of any stressed word should be such that it is suitable for being the last or one but last word in an utterance. This is contrary to what we know about the temporal structure of speech: a word spoken in isolation and the same word at the end of an utterance have approximately the same duration. (See for example relevant data in De Rooij, 1979). The word duration gets systematically shorter as more material follows in the utterance, it does not get systematically longer, compared to the duration in isolation, as more material precedes it. The anticipatory shortening effect can be illustrated with some data from De Rooij (1979), who measured durations of sequences of a vowel plus following plosive silent interval in the Dutch stressed monosyllabic words [pe:t] and [Xa:t], as a function of increasing number of following words in the sentence. Durations, for the present purpose averaged over twelve different cases, were as follows:

Table 1. Average durations of vowel plus silent interval in Dutch stressed monosyllable words, for 0 - 5 following words in the same utterance. N = 12. After De Rooij (1979).

0 following words:	242 ms
1 following word:	216 ms
2 following words:	186 ms
3 following words:	176 ms
4 following words:	167 ms
5 following words:	167 ms

That a duration of 167 ms would not be suitable when no words follow in the utterance, can easily be confirmed in a simple class room demonstration: an utterance initial word, artificially isolated from the utterance by gating, sounds unacceptably short. An utterance final word presented in isolation sounds alright. This demonstration also works after intonational cues have been made identical by means of LPC-analysis, manipulation of pitch, and resynthesis. De Rooij's data suggest that his speakers had a lookahead of three or four words. This may not be surprising: these data were obtained with read out sentences, where lookahead was prepared for the speakers in print. To my knowledge no such or similar data exist for spontaneous speech. If in future research similar effects are found in spontaneous speech, these effects can be interpreted in terms of lookahead during speech production.

Summing up: Anticipatory shortening is really what the term says. It is not persevatory lengthening, as suggested by Levelt. Therefore, anticipatory shortening can be used as an indication of the amount of lookahead in speech production. In reading out sentences, lookahead appears to be in the order of three or four words. Data on spontaneous speech are still lacking.

5. Anticipatory errors of speech

When someone says: "sil.. filter cigarette" (example taken from Hockett, 1967), we assume that the [s] of "cigarette" is anticipated inappropriately in the pronunciation of the word "filter", replacing the [f] of "filter". This can only be explained by assuming a lookahead of at least one word. By the same token a slip like "knife with the salami" instead of "salami with the knife" seems to suggest a lookahead of at least three words. From Lashley (1951) onwards many students of speech errors have argued in these or similar terms that anticipatory errors show "that speakers must have access to a representation that spans more than the next word of the utterance" (Shattuck-Hufnagel, 1979). Of course, as errors of speech themselves by their very nature are the exception rather than the rule, they do not tell us what the amount of lookahead normally is. They can only provide hints about what the lookahead sometimes can be.

In order to discuss errors of speech in terms of lookahead, we have to distinguish between 'origin' and 'target'. 'Origin' is the position where a particular entity belongs in the error free version of the utterance. 'Target' is the position where this entity ends up in the speech error. So in the lapse quoted from Hockett, the origin is the first phoneme of "cigarette", and the target is the first phoneme of "filter". We assume that the distance between target and origin in the intended utterance provides a clue as to the amount of lookahead at the time the error was produced. As we cannot be certain that speech errors involving phonemes as misplaced entities and speech errors involving morphemes or words as entities are generated at the same stage of speech programming, the two classes of errors are best kept separate.

Many years ago I made some counts of distances between targets and origins in a collection of Dutch and German errors of speech (Cf. Nooteboom and Cohen, 1975). For phonological errors distances were expressed in syllables. This gave the following numbers (N = 1057):

Table 2. Distances in syllables between origin and target in phonological anticipatory errors of speech in Dutch and German. The syllable containing the origin is, and the syllable containing the target is not counted. N = 1057.

1 syll. lookahead:	34 %
2 syll. lookahead:	29 %
3 syll. lookahead:	16 %
4 syll. lookahead:	10 %
5 syll. lookahead:	3 %
> 5 syll. lookahead:	8 %

Recalculating these numbers on the basis of an average word length in of 1.8 syllables (an estimate obtained from the spontaneous utterances in the same collection of speech errors), gives the following estimates of lookahead in numbers of words:

Table 3. Distances in words in phonological anticipatory errors of speech in Dutch and German. The word containing the origin is, and the word containing the target is not counted. N = 1057.

0 words lookahead:	c. 15 %
1 word lookahead:	c. 50 %
2 words lookahead:	c. 23 %
> 2 words lookahead:	c. 12 %

In order to check whether these estimates are at all realistic, I counted lookahead spans expressed in words in the selection of errors published in Fromkin (1973), assuming that the selection was random with respect to material spans involved, and skipping all non-anticipatory errors and lexical errors and also some errors I found hard to interpret. I also excluded a list of within-word errors because these were obviously selected according to the material span involved. There remained 231 anticipatory phonological errors, of which 10 (4 %) showed zero lookahead, 129 (56 %) one word lookahead, 62 (27 %) two words lookahead, 22 (9 %) three words lookahead, and 9 (4 %) of four words lookahead. Given the limited size of this sample, these numbers come reassuringly close to the above estimates. Some examples of Fromkin's speech errors are:

Table 4. Examples of phonological errors of speech, taken from Fromkin (1973).

0 words lookahead:	significantly ? significlantly
1 word lookahead:	roman numeral ? noman numeral
2 words lookahead:	a Canadian from Toronto ? a Tanadian .....
3 words lookahead:	the hiring of minority faculty ? the firing of...
4 words lookahead:	Paris is the most beautiful city ? Baris ...

The estimates obtained from the Dutch/German corpus indicate that in 35 % of anticipatory phonolo gical speech errors lookahead is more than a single word. This is a considerable proportion, suggesting that lookahead of more than one word may not be all that exceptional.

Not only phonemes move around in speech errors. Morphemes and whole words also get misplaced. The following numbers were obtained in counting words between targets and origins for anticipatory speech errors involving lexical items (morphemes and words) as entities changing position (N = 147):

Table 5. Distances in words between orig in and target, in anticipatory lexical errors of speech in Dutch and German. The origin is, the target is not counted. N = 147.

0 words lookahead:	7 %
1 word lookahead:	34 %
2 words lookahead:	24 %
3 words lookahead:	22 %
4 words lookahead:	10 %
> 4 words lookahead:	3 %

Fifty-nine percent of these errors involve a necessary lookahead of more than one word! That is much more than the 35 % we estimated for phonological speech errors. Obviously speakers look farther ahead when selecting and ordering lexical items than when spelling out the selected lexical items as strings of ordered phonemes. This seems to provide an answer to a question raised by Shattuck-Hufnagel (1979, p. 329): "Does the size of the span change during the planning process; e.g. is it longer when syntactic structure is being computed, shorter when phonological details are being worked out?" But whether the span changes or not depends on how we count the entities in the span. I will shortly come back to this.

One way of interpreting the difference between the two classes of speech errors is to assume that they reflect two different stages of mental programming. One stage, generating the surface structure, is concerned with selecting and ordering lexical items, and one stage deals with spelling out phonological forms and setting up phonetic plans for speaking these items in coherent stretches of speech. This interpretation is in line with Levelt (1989), who discusses speech errors involving exchanges of words and morphemes in his chapter on the generation of surface structure, and discusses phonological errors in his chapter on phonetic plans for words. Levelt also points out that misplaced lexical items attract the pitch accent, case marking and inflectional forms that go with their new position:

(a) "the knife with the SALAMI" instead of "the salami with the KNIFE"

(b) "Bis er es bei Dir abholt" instead of "Bis Du es bei ihm abholt"

(c) "Dat is nieuwer dan een dure" instead of "Dat is duurder dan een nieuwe"

In (a) "salami" gets the pitch accent that "knife" should have had in the intended utterance. In (b) the German pronouns for second and first person not only swap positions, but then receive the grammatically correct case markings going with their new positions. In (c) the Dutch content morphemes "nieuw" and "duur" exchange position, and then the comparative suffix changes correctly from "-der" to "-er", adapting itself to the incorrectly placed content morpheme. There are many such examples in the literature, showing that errors of this type take place during grammatical encoding rather than during phonological encoding. Errors of type (b) and (c) show that function words and inflectional morphemes are not yet spelled out phonologically at the moment the speech error is generated. Presumably they are present as abstract syntactic and/or semantic functions or labels, that are about to be attached to appropriately or inappropriately selected content morphemes, and only thereafter receive their phonological form. For the present purpose this means that it is hard to know how to interpret the quantitative data given earlier in terms of lookahead. Do we count formless abstract function words as words, and formless abstract inflections as morphemes? Or do we only count content morphemes in estimating the lookahead? If we do the latter, the distribution of amounts of lookahead for lexical errors becomes much more similar to the one for phonological errors, showing a majority of cases with only one item lookahead. Such items are then, of course, much more grammatically complex than the phonological forms in the phonetic plan. However, there does not seem to be a principled way to decide what should and what should not count as ordered entities at the stage of grammatical encoding.

Summarizing: Phonological speech errors may be used as a source of information about the amount of lookahead at the stage of phonological encoding, generating a phonetic plan. As in an estimated 35 % of such errors lookahead is at least two words, we may conclude that lookahead of more than a single word is not exceptional. Lexical speech errors on the other hand reflect planning at the stage of grammatical encoding. In terms of the utterance to be produced, speakers look farther ahead at this stage than at the stage of phonological encoding, three or four words lookahead not being exceptional. It is difficult, however, to decide on the nature and number of ordered entities actually involved at this stage of programming.

6. Discussion

We have no direct access to the size of the phonetic plan underlying speech production. Quite literally, we do not know what we do we speak. Estimates of the extent of preprogramming during speech production can only come from indirect evidence, such as acoustic/phonetic aspects of speech depending on what is yet to come, and recorded slips of the tongue. We have seen that lexical speech errors, such as "knife with the salami", cannot be taken to reflect spans of attention on the level of the phonetic plan. They rather betray lookahead during grammatical encoding, often comprising three of four words, and sometimes more. It seems to me that lookahead during grammatical encoding in spontaneous speech production could in principle also be investigated by looking at material spans over which selection dependencies are maintained, for example dependencies between early auxiliary verbs and later past participles in Dutch. My main concern here, however, is not with grammatical but with phonological encoding.

The brief review of some available empirical evidence given above suggests that, although one word lookahead may be sufficient for the production of fluent speech, a lookahead of more than a single word is far from exceptional in spontaneous speech production. Often lookahead is two words and occasionally lookahead may be three or four words.The strongest evidence to this effect stems from phonological speech errors. There is no way to know, of course, whether these estimates are biased: It may be that the probability of speech errors increases with the amount of lookahead. If so, the frequencies of occurrence of particular material spans over which errors occur would not reflect frequencies of occurrence of amounts of lookahead in error-free speech production. On the other hand, the material spans counted contain what is minimally necessary to explain the errors concerned. There is no way of knowing whether actual lookahead is generally more than this. Because of such uncertainties it is worthwhile to look at empirical evidence from different sources.

As we have seen, anticipatory shortening of stressed syllables in sentence production can be used as evidence for the amount of lookahead during speech production. Available evidence suggests a lookahead of three or four words during reading aloud. Unfortunately there are no available data on spontaneous speech. This is, of course, not accidental. Speech sound durations are strongly affected by a great many factors that also show strong interactions (Klatt, 1976; Nooteboom, 1991; Van Santen, 1992). In order to get a precise estimate of how many upcoming words still have an effect on the shortening of an utterance initial stressed word or syllable, one needs rather precise control over the speech material. Spontaneous speech yet to be produced, by its very nature does not easily allow us such control. But perhaps in the future this lack of control can be compensated for by using large speech data bases and quantitative techniques such as proposed and tested on read out texts by Van Santen (1992). Further data on anticipatory shortening as a measure of lookahead would be particularly interesting because it seems to be virtually always present and would thus potentially provide a relatively rich source of empirical evidence.

We have also seen that the material spans covering rise-fall configurations or hat patterns in Dutch speech melodies betray the amount of lookahead minimally present at the moment the rise is produced. Of course, this situation is hardly representative of what at other moments happens in speech production: Hat patterns over more than a single word are not always allowed and are never obligatory. In many utterances speakers do not have the option to produce a hat pattern over more words, for example because there is only one accented word in the utterance. In most other utterances there is only one point where the speaker has an option to produce a hat pattern over more words. Of all cases where speakers have this option, they decline taking it in 46 % of the cases, possibly because they lack the necessary lookahead at that moment, or for other unknown reasons. It would be interesting to know whether the accented words at which speakers decline to take the option of an extended hat pattern do or do not show anticipatory shortening as a function of the number of upcoming words in the utterance. As it is, the data on hat patterns at least show that a lookahead of more than a single word is not something very special, and that lookahead of three or more words does occur during normal spontaneous speech production.

We started our search for empirical evidence for lookahead with the starting frequency and slope of declination. In principle this could be a fairly reliable indicator of planning ahead in speech production. Although pitch is influenced by the segmental structure of utterances, the influence is minor as compared to what happens to speech sound durations. We do not need very precise control over the speech material in order to test hypotheses as to the correlation between the length of intonational phrases and the starting frequency and slope of declination. A large data base of spontaneous utterances, preferably with many isolated utterances from a single speaker, or from each of a few speakers, should allow us to asses how many upcoming words affect the course of pitch at the onset of the utterance. Data on reading aloud show that there is preprogramming involved. Data on spontaneous speech suggest that there also preprogramming is not excluded. Statistical studies specifically directed at the amount of preprogramming are still lacking, though, both for reading aloud and for spontaneous speech. My prediction is that such studies will show that a lookahead of three or four words in spontaneous speech is far from exceptional.

References:

Borden, G.L., K.S. and Harris (1980). *Speech science primer. Physiology, acoustics, and perception of speech*. London/Baltimore: Williams and Wilkins.

Cohen, A. and J. 't Hart (1967). The anatomy of intonation. *Lingua 19*, 177-192.

Collier, R. (1972). *From pitch to intonation*. Diss. Louvain.

Cooper, W. E. and J.M. Sorensen (1981). *Fundamental frequency in sentence production*. Berlin/Heidelberg/New York: Springer Verlag.

De Pijper, J.R. (1983). *Modelling British English intonation*. Diss. Utrecht.

De Rooij, J.J. (1979). *Speech punctuation. An acoustic and perceptual study of some aspects of speeh prosody in Dutch*. Diss. Utrecht.

Fromkin, V.A. ed. (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Hadamard, J. (1945). *The psychology of invention in the mathematical field*. Princeton: Princeton University Press.

Hart, J.'t, Collier, R. and Cohen, A. (1990) *A perceptual study of intonation. An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Hockett, C.F. (1967). Where the tongue slips, there slip I. In: *To Honor Roman Jakobson*. The Hague: Mouton. Also in Fromkin (ed.) *Speech errors as linguistic evidence* (pp. 93-119). Mouton, The Hague: Mouton.

Klatt, D.H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59*, 1208-1221.

Lashley, K.S. (1951). The problem of serial order in behaviour. In: L.A. Jefferess (ed.) *Cerebral mechanisms in behavior: The Hixon symposium.* New York: Wiley.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge/London: MIT Press

Lindblom, B.E.F. (1968). *Temporal Organization of Syllable Production.* Speech Transmission Laboratory, Quarterly Progress Report No 2-3/1968 (pp. 1-5). Stockholm: Royal Institute of Technology.

Nakatani, L.H., O'Connor, J.D. and Aston, C.H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica, 38,* 84-106.

Nooteboom, S.G. (1972). *Production and perception of vowel duration*. Diss. Utrecht.

Nooteboom, S.G. (1991). Some observations on the temporal organisation and rhythm of speech. *Proceedings of the XIIth international congress of phonetic sciences, Aix- en-Provence, August 19-24, 1991, vol. 1,* pp 228-237.

Nooteboom, S.G. and A. Cohen (1975). Anticipation in speech production and its implications for perception. In: A. Cohen and S.G. Nooteboom (eds.) *Structure and process in speech perception*. Springer-Verlag Berlin/Heidelberg/New York, pp. 124-142.

Ohala, J.J. (1978). Production of tone. In: V. Fromkin (ed.) *Tone: A linguistic survey. New York:* Academic press.

Penrose, R. (1989). *The emperors new mind*. Oxford: Oxford University Press.

Santen, J.P.H. van (1992). Contextual Effects on Vowel Duration. *Speech Communication, 11,* 513-546.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In: W.E. Cooper and E.C.T. Walker (eds.) *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295-342). Hillsdale, New Jersey: Lawrence Erlbaum ass., publishers.

Willems, N. J. (1983). Towards an objective course in English intonation: standardized precepts. In: M. van den Broecke, V. Van Heuven, W. Zonneveld (eds.) *Sound structures, studies for Antonie Cohen.* (pp. 281-296). Dordrecht/Cinnaminson: Foris publications.