Developing second-language listening comprehension: Effects of training lower-order skills versus higher-order strategy.

Poelmans, P.

Citation for published version (APA):
Chapter 2  

Background

2.1. Introduction

In the field of second language learning four basic language skills are distinguished. These are ordered along two dimensions:

(i) **modality**, that is the difference between the auditory language mode versus the visual mode, and

(ii) **processing activity**, that is the process of either encoding or decoding.

Together these two binary dimensions define the so-called four skills, as indicated in Table 2.1.

*Table 2.1: The four language skills.*

<table>
<thead>
<tr>
<th>Modality</th>
<th>Processing activity</th>
<th>encoding</th>
<th>decoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>auditory</td>
<td>speaking</td>
<td>listening</td>
<td></td>
</tr>
<tr>
<td>visual</td>
<td>writing</td>
<td>reading</td>
<td></td>
</tr>
</tbody>
</table>

The listening skill is defined, then, as the process by which the language user decodes auditory input, i.e., speech. This use of the term ‘listening’ deviates from its ordinary use, in which listening simply refers to the act of intentional hearing, in much the same way as looking refers to the act of intentional seeing.

In other disciplines the listening skill is often called listening comprehension, which is the complex of processes that transform an auditory stimulus to a mental reconstruction on the part of the listener of the speaker’s intention. In this thesis we will use the terms ‘listening (skill)’ and ‘listening comprehension’ indiscriminately.
The aim of this chapter is to provide the reader such theoretical background information on listening and language learning as is needed to follow the exposition in the later chapters.

2.2. The listening process

To communicate adequately it is important for a listener to understand what the speaker says. Understanding spoken language can be described as an inferential process based on the perception of several cues rather than a simple match between sounds and meaning. The listening comprehension process is a combination of (roughly) four sub-processes or modules:

(i) **hearing**: the auditory reception of an acoustic signal (perception), as is also performed on non-speech sounds,

(ii) **categorisation of sounds**: categorising incoming sounds in terms of the sound categories of the language

(iii) **word recognition**: breaking up the stream of sounds into linguistic units (morphemes, words) and retrieving their meaning from long term memory, and

(iv) **comprehension**: integrating the meanings of the words in their sequence into an interpretation of the entire utterance, i.e., a reconstruction of the speaker’s communicative intention (= message).

Obviously, the third and fourth modules draw heavily on linguistic knowledge. Both lexical knowledge, needed to recognize words, and knowledge of the rule system of the language, needed to decode the grammatical relationships among the words that make up the sentence, are part of the listener’s linguistic competence (Ellis 2000). However, especially at the third stage, knowledge of the world, i.e., non-linguistic knowledge, plays an often-indispensable role in the listening process.

Figure 2.1 shows the different sub-processes of the listening comprehension process. Each of the sub-processes displayed in Figure 2.1 can be a source of listening comprehension problems (see section 2.3.).
Figure 2.1: Simplified overview of the sub-processes of the listening comprehension process.

Before embarking on a discussion of the modules included in Figure 2.1, it is expedient to introduce a parameter that is often used with reference to psycholinguistic processing models, viz. the direction of the information flow in terms of bottom-up versus top-down. In Figure 2.1 sensory information enters the system at the periphery, which is popularly called the bottom-end of the system. In successive stages the lower-order information is transformed to more abstract, higher-order, representations. Auditory details are lost in the transformation process, each time when a higher-order representation is derived. The loss of detail is probably a prerequisite if the brain has to keep a longer stretch of speech in working memory. The most abstract representation, generated by the most central module, is the overall interpretation of the discourse, the gist of the message, in which details of individual sounds, words and grammatical structures are absent.

In addition to the bottom-up flow of information, i.e., from sensory periphery to central processing by the brain, there is a top-down information flow from the brain down to the periphery. Top-down information flow may occur between modules. For instance, when a (predictable) sound in a spoken word is (electronically) replaced by a noise burst (or cough), the listener will recognize the word as if no sound was missing; in fact, the listener will not be able to tell the experimenter which sound exactly was missing from the word. The word was recognized from partial sensory input, and the missing phoneme was restored (hence: ‘phoneme restoration effect’, cf. Obusek & Warren 1973; Samuel 1990) on the basis of the complete lexical specification. Information available at a higher, more central, representation is fed back, then, to the lower level. Top-down information is typically used to limit the number of choices that has to be made by a lower-level module, which serves to free processing capacity in working memory. When a process, or module contained by it, crucially depends on both bottom-up and top-down information streams, it is called ‘interactive’; when the process uses bottom-up information only, it is called ‘autonomous’.
As is indicated by the bi-directional arrows in Figure 2.1, we take the view that the modules in the listening process are typically interactive; only the peripheral hearing process is considered to be autonomous.

The first sub-process of the listening skill is **hearing**. We define hearing as the relatively peripheral process of filtering the relevant speech signal from the multitude of acoustic events that simultaneously strike the ear. Even under optimal listening conditions the speaker’s voice must be isolated from background noise. In more complicated listening situations, such as may occur when several speakers talk at the same time, in a noisy environment, the process by which a listener is able to attend to just one speaker’s voice is ill understood. For the purpose of the present study we will assume that the sub-process of filtering the relevant signal from the complex auditory input is language independent, and never poses a specific problem for the foreign language learner.¹

The linguistic processing of the output of the hearing stage is the focus of this thesis. As a first language-specific transformation listeners **categorise** incoming sounds in terms of the sound categories of their language, this is the second sub-process of listening. Speech sounds differ along a limited number of phonetic dimensions, each of which may depend on several auditory cues. For instance, whether a consonant is perceived as an intervocalic voiced /b/ or voiceless /p/ – as in the minimal pair **ruby ~ rupee** – depends on a multiplicity of auditory cues, such as the duration and intensity of a noise burst, the duration of a (nearly) silent interval separating the plosive from the preceding vowel, the duration of the preceding vowel, the abruptness of the amplitude and formant frequency changes at the edges between the consonant and the abutting vowels, and the course of the fundamental frequency at the onset of the post-consonantal vowel (Hayward 2000: 117-121, Rietveld & van Heuven 2001: 62-64). It is commonly held

¹ There is at least one aspect of hearing that demonstrably involves the use of top-down information, i.e., speaker separation on the basis of prosodic continuity. It has been shown (Darwin, 1975 in Nooteboom, 1996: 669-670) that a listener is able to attend to one speaker’s voice in the midst of competing noises and competing (even louder) speakers. This is the so-called cocktail-party effect described by Cherry (1953). One very important cue the listener uses to track one speaker’s voice is the predictability of the pitch pattern; the listener will attend to the speaker whose voice continues on the pitch level that the listener expects from the preceding spoken context. Intonation patterns differ between languages. Therefore, generating predictions of the pitch level of the next syllable involves language-specific, top-down information. For expository reasons, we will nevertheless assume that hearing proceeds in a strictly bottom-up fashion.
that a dozen binary phonetic dimensions (theoretically yielding ca. $2^{12} = 4096$ different sound categories or ‘phonemes’, 15-75 in each language) suffice to capture the variety of vowels and consonants in the languages of the world. Each of these binary dimensions is coded in turn by a limited number of primitive acoustic cues of the type exemplified above. However, languages differ widely in the number and type of phonemes, and in the precise division of the multi-dimensional phonetic space over the sound categories that exist in the language.

Apart from the segmental categories that define the various vowels and consonants in the language, spoken utterances are characterised by so-called suprasegmental (or: prosodic) properties. These are properties that do not belong to specific, individual speech sounds but subdend larger units of at least the size of a syllable. Prosodic features serve to break up the continuous utterance into smaller chunks (clauses and phrases, sometimes inaptly referred to as thought units) by the presence of pauses and boundary-marking pitch changes, and to mark one syllable or word as the focus of the speaker’s attention within the chunk (accentuation) (Nooteboom 1997, and references given therein). Typically the suprasegmental properties guide the interpretation process whilst the segmental cues serve to access words in the mental lexicon (Cutler 2001).

The third sub-process of the listening skill is **word recognition**. The process of word recognition has been the object of phonetic and psycholinguistic research for years; the extensive body of research has resulted in several word recognition models. A few well-known models are (in chronological order) the Logogen model (Morton 1969), the Cohort model (Marslen-Wilson & Welsh 1978), the TRACE model (McClelland & Elman 1986), the Shortlist model (Norris 1994) and the Merge model (Norris et al 2000). Parameters in which the models differ are, firstly, the assumptions about the nature of the elements that make contact with the mental lexicon (e.g. phonemes in the Logogen model but sound features in the TRACE model) and, secondly, the presence or absence of a feedback possibility (top-down flow); the interactive TRACE model allows both bottom-up and top-down information streams whereas the Shortlist model is of the autonomous type and allows bottom-up information only.

The models mentioned here are based on the word recognition process in the native language. The word recognition process in a non-native language has to deal with the incongruence between native and non-native language elements. It is only recently that researchers have begun to work on an explicit non-native word recognition model. So far this research resulted in the BIA (Bilingual Interactive Activation) model for visual non-native word recognition (Dijkstra & van Heuven 1998; Dijkstra, van Heuven, &
Grainger 1998) and the BIMOLA (Bilingual Model of Lexical access) model, which is still under development, for non-native spoken word recognition (Grosjean 1988, 1997).

The fourth sub-process of the listening skill is the comprehension (and/or interpretation) of the stream of sounds that is uttered by the speaker. When the incoming words are recognized, they are assigned to grammatical categories (various classes of content words and function words) and the structural and semantic relationships between the words are established: the input gets parsed. Because of the redundancy in communication it is not necessary for the listener to make a precise and exhaustive grammatical analysis; moreover, the rate of delivery and the limited processing capacities of the listener make it rather impossible for language users to construct detailed grammatical analyses. The parsing process critically depends on the learner’s knowledge of the language.

That knowledge of the language influences the comprehension process seems obvious; it is only when language users know words and grammatical structures of a language that they can recognise them.

The comprehension sub-process does not only depend on lexical and grammatical, but also pragmatic knowledge and ‘knowledge of the world’.

As far as pragmatics is concerned, there are three key notions that play a role here (Rost 2002):

(i) **Deixis.** Deictic elements are often employed by a speaker, and interpreted by the listener, as elements that do not directly – but only indirectly – refer to objects in the ‘real world’. Consider the following discourse: *Beatrix is the queen of The Netherlands. She lives in The Hague.* Here the proper noun *Beatrix* holds a direct and constant correspondence to a person in the real world. The interpretation of the deictic element *she* in the second sentence is indirect and variable. To resolve the referent of *she*, the listener first has to establish that *she* is co-referential with *Beatrix*. Deictic elements are commonly used for making variable, context-dependent reference to, for example, moments in time (e.g., *yesterday, now*), locations in space (e.g., *here, there*), and objects in the time-space continuum (through pronouns).

(ii) **Intention.** The notion of intention refers to the speakers’ goals when they produce an utterance; they want to have some kind of influence on the listener, and act in a certain way to accomplish the desired influence. Generally the speaker’s intention is completely determining the choice of words, the grammatical structures used
and sometimes also the gestures. There are situations, however, where an inferential leap is required on the part of the listener in order to correctly interpret the speaker’s intention. Irony or sarcasm typically has to be inferred from the context. Or, to give one more example of inferential work that may have to be done by the listener, a speaker saying ‘It is rather cold in here’ should actually be interpreted as issuing a subtle command for the addressee to get up and close the window (Searle 1975).

(iii) **Strategic use.** Speakers generally have multiple options at their disposal to accomplish their goal. It is an integral part of the comprehension sub-process that the listener does not only recognize what it is that the speaker wishes to accomplish (i.e., his intention) but that he is also aware of the way the speaker uses to reach this objective, for instance by being authoritarian, by expressing anger, by being friendly or polite, etc. The so-called indirect speech act illustrated under (ii) above is a clear example of a speaker-strategy that requires a lot of inferential work on the part of the listener.

**Knowledge of the world** does also play a role in the comprehension and interpretation of speech. For a good interpretation of direct speech, and certainly for the interpretation of indirect speech, it is important that the listener and the speaker share background knowledge. Effective recovery of background knowledge and use of the pragmatic keys described above will optimise the listening process. Notions that are linked with the activation of the listeners’ background knowledge are ‘schema’ and ‘script’.

Schemata are general cognitive constructs of knowledge that listeners use to model current events and situations and thereby bring them into alignment with familiar patterns of experience (Widdowson 1983). Listeners use schemata to order and categorise information that makes it possible to interpret and remember this information. A more detailed cognitive framework than a schema is called a script. Scripts, then, can be conceived of as specific modules within a larger schema. A script is typically procedural and comprises the time-course of an event. A well-known example illustrating the notions schema and script is the act of having dinner in a restaurant. Most listeners have some notion of events that typically happen in a restaurant; the knowledge of the environment (tables, napkins, food, check) can be marked as a schema, the knowledge about the order of an event (waiter asks what one wants to eat, after the meal he or she will give the check) is a script.

Every time a language user is engaged in reading or listening, existing schemata and scripts are updated or new ones are made. The many
schemata and scripts adults have are interrelated and their level of activation is influenced by, for example, the frequency and the recency of their use.

Bransford and colleagues explored the importance of prior knowledge and the processes of remembering and comprehension. Bransford and Johnson (1973, in Harley 1998), for example, showed that ideas that were presented in a read-aloud text, were better recalled when context was given before the reading, in other words when the correct schema and scripts had already been activated. They also showed that remembering the ideas of a text is facilitated when the context is very familiar to the listeners. Here we assume that better recall was caused by better comprehension. More direct evidence for the importance of activating scripts and schemata was provided by Sulin and Dooling (1974, in Harley 1998), who showed that activated background knowledge may interfere with the comprehension of an utterance; it can make listeners remember ideas that were not presented. Schemata and scripts enable listeners to comprehend references to people and events within the cognitive framework that were not explicitly mentioned by the speaker. This way the communication between speaker and listener is made more efficient, accelerating and facilitating the comprehension process.

2.3. The L2 listening process

There are obvious differences between acquiring listening comprehension skills in the first and in the second language. The acquisition of the L1 listening comprehension skill happens largely at the same time that the child develops its general cognitive abilities. The development of the L2 listening processes occurs after the learner’s cognitive development is more or less completed. When acquiring a foreign language, learners have to learn mainly what the L2-words and grammatical forms refer to (Gathercole & Thorn 1998); they have to acquire the differences between their L1 and the L2. In this light the idea of ‘conceptual redeployment’ has to be mentioned (Churchland 1999).

Conceptual redeployment can be defined as the process in which a fully developed conceptual framework that is in regular use in some domain of experience or comprehension, for example L1 comprehension, comes to be used for the first time in a new domain, for example in L2 comprehension. In the case of L2 learning, adults must reorganize semantic knowledge of words and concepts into a new domain of language use,

---

2 This of course does not hold for bilingual children. The processes described here are based on L2 learning in a monolingual environment.
namely in their L2. The L2 learners already have cognitive frameworks but these have to be filled out with new (L2) language forms. The fact that new references (of known concepts) have to be learned and the change of conceptual frameworks while the cognitive development of the learner has already been developed makes learning a second language a more conscious process than L1 acquisition.

For successful second language learning it is important that there is enough correct language input. The kind of input that second language learners receive has been the object of research for several years. The speech input directed by L1 speakers towards L2 learners, which is often referred to as ‘foreigner talk’ shows some similarities with the modified input of child-directed speech; adjustments are noted at several linguistic levels. For example the phonology and the phonetics of foreigner talk differ from regular L1 talk in that there is more use of pauses, that there are fewer contractions and that there is a wider pitch range.

The discussion on the effect of this modified input and the way L2-learners deal with non-modified input has not been ended yet. In L2-research four different approaches of dealing with input can be distinguished: (i) studies that focus on the frequency of linguistic structures in the input, (ii) studies that focus on discourse and the way discourse construction is linked to acquisition, (iii) studies that relate the input to the output of L2-learners, and (iv) studies that focus on the comprehensibility of the input (see Ellis 2000 for a detailed description of these kind of studies).

A theory about the relation between input and improving comprehension is the ‘Input Hypothesis’ (Krashen 1982, 1985). The Input Hypothesis says that learners naturally develop their understanding and comprehension of the language by understanding only input that is slightly above their current language proficiency level.\(^3\) If the current level of the language learner is denoted as \(i\); then the next-higher level \(i+1\) can be reached by training with understanding and comprehension input at level \(i+1\). Input materials at the current level \(i\) contribute no new information, whereas input at level \(i+2\) would be too difficult to understand at all. Lack of

---

\(^3\) Ellis (2000) gives an overview of similarities and differences between child-directed speech and foreigner talk.

\(^4\) The Input Hypothesis further claims that comprehensible input is not the only requirement for language learning: learners need to be willing to ‘let the input in’ they comprehend. The hypothesis claims that fine-tuning is not necessary; simplification and the help of extra-linguistic information and context make the input comprehensible. A last point of the theory is that learner production does not contribute to acquisition. (see Ellis 1994 for a discussion of studies for and against the Input Hypothesis)
empirical support (and even of testability) of this theory is a source of criticism.

As was made clear in the previous paragraphs, acquiring listening comprehension in an L2 is a different cognitive process than the acquisition of listening skills in one’s mother tongue. When learning a second language, learners already know how to listen in their L1; listeners have (unconscious) knowledge of (L1) listening strategies and their effects on the listening comprehension process. They can use some of these strategies, schemata and scripts while listening to the L2. In fact, for beginning language learners this higher-order knowledge is the most important source they have. By leaning on these mainly higher-order strategies, there will be less focus on the exact linguistic features of the utterance, i.e., less focus on the input itself. It is thus not the case that all input contributes to the learner’s knowledge and understanding of the language; not all input becomes ‘intake’. Sharwood Smith (1986) makes a distinction between input that helps the learner interpret the utterance and input that learners use to expand their language knowledge. Only the latter kind of input may become intake. Ellis (2000) defines intake as ‘information stored in temporary memory which may or may not be subsequently accommodated in the interlanguage system’. Chaudron (1985, in Ellis 2000) describes intake as a process that mediates between target language input and the learner’s internalised set of rules.

As was made clear, language learners may lean on higher-order processes as well as on lower-order processes while processing language input. What now is the importance of these two kinds of processing in language learning? Research has been done to establish the relative importance of lower-order versus higher-order processes, and to determine the relationship between these processes and overall listening proficiency. Tsui and Fullilove (1998, p. 433) state that: ‘While some studies have found that less-skilled readers/listeners are deficient in top-down processing skills, others have contradicted this, citing evidence that, in fact, less-skilled readers/listeners lack bottom-up processing skills.’ The authors looked at the results on (part of) The Hong Kong Certificate of Education Examination (HKCEE) English Language, in order to establish the importance of bottom-up versus top-down processes as a discriminator of listening performance, and this way found out what kind of processes are typically used by skilled listeners. The material had the design of the paradigm often used in listening research and education, namely a listening text with multiple-choice questions. There were furthermore, two kinds of schemata: the first type of schema activated by the initial input was consistent with the following input; in the second type, the schema activated in the beginning was not consistent with the subsequent input. In this latter case, participants had to be able to revise the schema in order to give a correct answer; here the bottom-up
processes are important. In the case of the ‘matching’ schemata, participants can rely largely on top-down processes to give a correct answer. The authors concluded that bottom-up processing was more important than top-down processing in discriminating the listening performance. They state that ‘... on the one hand because less skilled L2 listeners are weak in bottom-up processing, initially they need plenty of contextual support to compensate for the lack of automatised linguistic decoding skills. On the other hand, they need to learn to become less reliant on guessing from contextual or prior knowledge and more reliant on rapid and accurate decoding of the linguistic input.’ (p. 449). In other words, as the decoding (or recognition) process becomes more automatised, the reliance on the higher-order processes may become weaker. This indicates that skilled learners make use of both top-down processes and (automatised) bottom-up processes, less skilled learners lack the automatisation of the lower-order processes.

In conclusion we can state that both bottom-up and top-down processes are important for successful listening comprehension. In second language learning, the difference between skilled and less-skilled learners lies in the status of the lower-order processes (ranging from more controlled to fully automatised). This implies that to improve the listening comprehension skills of second language learners, the focus has to be on lower-order processes.

2.4. Problems in listening

Listening problems (that result in miscommunication) may have several different sources; the problem can be pure physical, for example a damage of the inner ear, the problem can be linguistic, for example a problem with the recognition of speech sounds, or the problem can be of a more abstract cognitive level, for example with the interpretation of the message.

Let us now describe in more detail the problems L2-learners may have in executing some specific sub-processes of listening comprehension. A first sub-process that can be a source of problems is the identification of L2-phonemes. One popular view is that native speakers identify speech sounds due to the so-called perceptual magnet effect, which means that they recognize sounds by comparing them to prototypical phonemes (Rost 2002). These prototypes lie within a range of allowable variation; as long as the characteristics of the input sound lie within this range, the sound can be categorized correctly. Lack of knowledge of the L2 prototypes that leads to wrong categorisation, could be a first problem for second language learners in processing spoken language; if they do not have enough knowledge of the L2-specific phonemes, they will not be able to categorise the input sounds
correctly. Clear examples are the difficulties Chinese listeners have with the categorization of (English or Dutch) /l/ versus /r/. These two sound categories are not contrastive in Chinese as they are in English (e.g. lead ~ read). Since Chinese people do not observe the contrast in their mother tongue (they hear an unspecified liquid), it is only possible for them to recognize these sounds after intensive practice.

Of course, not all L2 sounds will cause problems. A distinction can be made between three different classes of L2 speech sounds, namely (Flege 1987):

(i) **Identical sounds.** These sounds of the L2 are the same as the sounds of the L1; L2 learners have nothing to learn. An example of an identical sound in Dutch and English is the lax high-mid front vowel /ɪ/.

(ii) **Similar sounds.** These sounds are not the same in L1 and L2; in terms of IPA transcription they share the same base symbol and differ in a diacritic only. An example is the /s/ in Dutch and English, which has intensity above 5 KHz in English (having a sharp timbre) but not in Dutch (leading to a duller timbre causing confusion with English /ʃ/). Subtly incorrect pronunciation of similar sounds will always give the L2 learner away as a non-native; it is believed that even long exposure and intensive training is of no avail here.

(iii) **New sounds.** These sounds show a substantial difference between the languages. At first, learners do not know that they have to pronounce a sound that they do not have in their native language. After a period of time (and intensive practice), the sound will become established as an extra category. An example is the English /æ/ like in hat ‘hoed’ which is a new category that has to be established by Dutch learners in between Dutch /e/ as in pret ‘fun’ and Dutch /a/ as is praat ‘talk’. New sounds are a source of difficulties that L2-learners experience, which may be overcome by setting up new, or redefining existing, prototypes and ranges of allowable variation for them. This may come about only after prolonged exposure to the L2 and/or explicit feedback.

---

5 The split of the undifferentiated liquid category into /l/ versus /r/ as is required of Chinese learners of English, would be yet another example of setting up ‘new’ sound categories.
Flege (1987) states that language users learn to categorize speech sounds in terms of phonetic categories of their L1 before the age of seven (in fact, they begin at birth – or even earlier). When the categorization is completed, new sounds (for example L2 sounds) will preferably be categorized in one of the existing and relevant L1 categories. Young children are still able to ‘create’ new categories for the L2 sounds whereas older children and adults are not.

A second sub-process that makes listening comprehension so difficult is the recognition of the spoken words. A characteristic of continuous speech that makes listening comprehension so difficult is that speech contains no clear auditory equivalent to the inter-word white spaces that we find in written text. The lack of clear word boundary markers in languages such as Dutch makes it hard to determine word beginnings and endings (the segmentation problem). It is important to realize that sounds in continuous speech are fluently coarticulated and assimilated, not only within words but also across word boundaries. These so-called Sandhi processes are only blocked at higher-order linguistic boundaries, at the level of the Intonation Phrase boundary (and higher) in Dutch (Berendsen 1986). The lack of explicit word boundary markers and the distortions of word shapes due to assimilation and reduction are the reasons why words that are known by the language user when presented visually, are often not recognized when they are part of continuous speech. Although phenomena like Sandhi and co-articulation are efficient for the speaker, since they facilitate pronunciation, the lack of word-boundary markers can cause misunderstanding or even the complete breakdown of word recognition.

Problems can also be caused by a form-meaning mismatch. If a listener identifies the form of a word correctly but has not enough knowledge of its meaning, the recognition process (and ultimately the communication process) will fail accordingly. Word-recognition problems, whether caused by faulty word-boundary identification or by insufficient lexical knowledge, are major sources of miscommunication in listening comprehension; this is, as is discussed, particularly the case in the second language listening comprehension process.

On the comprehension level researchers point out the effort for ‘complete understanding’ as a characteristic of L2 learners that causes problems.

---

6 Implicit word-boundary markings of a subtle nature are sometimes available in spoken language. For instance, there are languages that use stress to mark the beginning or end of a word. Hungarian, for example, has main stress on the first syllable of the word whereas Dari stresses the final syllable.

7 In this context form can be defined as an auditory presentation of the concept.
L2 learners often try to get a complete understanding of the utterance, anxious to miss any aspect of the message. A characteristic of a good L2-listener is the ability to distinguish between important and less important aspects of the utterance, as is also done in L1-listening. In languages such as Dutch and English prosody is used to signal the position of important parts of the sentence. Typically important words are pronounced more carefully and more slowly (some 10 percent, Nooteboom & Eefting, 1991; van Heuven 1998). Moreover, the prosodic head (stressed syllable) of important words is associated with a perceptually conspicuous change in vocal pitch (accentuation). Native listeners use these changes in speaking rate and pitch to predict the positions in the utterance where important words and syllables will be spoken. Experiments have revealed that (English) listeners attend more forcefully to bottom-up information at points in time where accented syllables are expected (Cutler, 1976; Cutler & Darwin, 1981). When the L2 listener is not familiar with the prosodic system of the target language the attention-focussing mechanism will not be employed. As a result the L2 listener will continually direct all his resources to bottom-up signal information, leaving less capacity for the recoding of the recognized units to a more abstract (semantic) representation. The importance of the division of work over the various modules in memory will be addressed in the next section. For now, we conclude that problems in L2-listening can be due to lower-order skills (recognition) as well as higher-order skills (comprehension).

In the two remaining sections of the present chapter we survey the literature on working memory and automatisation, respectively, in so far as these concepts are relevant to our study.

2.5. Memory

This section first briefly considers working memory (WM) as a component of memory in general and then focuses on the components that make up working memory, giving special attention to the so called phonological loop and the measurement of working memory capacity.

According to Carroll (1994), human memory can be divided into three units: the sensory stores, working memory, and long-term memory. Figure 2.2 illustrates the relations between the units (based on Carroll 1994).
The sensory stores take in the colours, tastes and tones we experience and retain these for a short period of time in an unanalysed form. The information of the sensory stores is refreshed and analysed in working memory. Long-term memory (LTM) is the unit that contains, amongst others, knowledge of the world. The functioning of working memory and the relation between WM and LTM is discussed below.

As Figure 2.2 shows, the input goes via the sensory stores and working memory to long-term memory. That WM is not just a gateway for information towards LTM, is indicated by the arrows from LTM to the other two units, they point out that knowledge that is stored in LTM may influence the processes of the sensory stores and WM. Evidence for LTM influence on the other units is given by for example Hulme, Maughan and Brown (1991). They showed that working memory span is better for words than for nonwords as it is easier to remember words than nonwords. Furthermore, Gathercole (1995) showed that memory span will be influenced positively not only by concrete concepts of words but also by knowledge about the structure of the language. She found that nonword recall was better for more wordlike nonwords than for less wordlike nonwords. Some theories therefore regard WM not as a unit separate from LTM but as that part of LTM which is currently in a state of activation beyond a certain threshold (Miyake & Shah 1999).

2.5.1. Working memory

The term working memory is used in several different ways in different areas of cognitive science (Miyake & Shah 1999 provide an in-depth
presentation and discussion of eleven WM models). In this present thesis the
definition of Baddeley is adopted: Working memory refers to a limited
capacity system of temporary storage and manipulation of input that is
necessary for complex tasks such as comprehension and reasoning (Miller et
al. 1960; Baddeley & Hitch 1974). The concept of WM is based on the
concept of Short-term memory (STM), the biggest difference between them
being a matter of dynamics. STM has mostly been considered as a passive
repository while WM has been attributed both a storage and a processing
function (Carroll 1994). Baddeley and Hitch (1974) proposed a WM model
that has been developed through the years to the most recent model shown in
Figure 2.3 (Baddeley 1986; Baddeley 2000). The three components of the
model presented are the visuo-spatial sketchpad, the episodic buffer, and the
phonological loop.

![Figure 2.3: The working memory model (Baddeley 2000).](image)

As can be seen in Figure 2.3, highest in the hierarchy of working memory is
the central executive, which is a supervisory controlling system. The central
executive is assumed to be modality free; it acts as a link between three
peripheral slave systems. The visuo-spatial sketchpad and the phonological
loop are modality specific; the visuo-spatial sketchpad deals with visual

---

*The largest difference between the revised model and the original memory model
is the focus on the processes of integrating information. The original model focused
on the isolation of the subsystems.*
information, while the phonological loop deals with speech based information. The episodic buffer deals with both visual and speech based information.

The episodic buffer is a new component in comparison with the Baddeley WM model of 1986. The need for a new storage system that is capable of dealing with all kinds of input was felt, amongst other things, by the lack of explanation for the recall of prose effect. This effect indicates that humans are able to recall 15-20 idea units, if these units are related. However, for the recall of unrelated words, failures occur once the number of words exceeds five or six. An explanation for this contrast is that in the case of related words, knowledge from LTM is involved. This indicates that information from various sources is used during the processing of related words: LTM, and, depending of the presentation mode of the items, phonological loop or visuo-spatial sketchpad. The combined information has to be stored in some kind of system. In the new WM model, the episodic buffer performs this back-up storage. Baddeley (2000) assumes the episodic buffer to be: ‘...a limited-capacity temporary storage system that is capable of integrating information from a variety of sources. It is assumed to be controlled by the central executive, which is capable of retrieving information from the store in the form of conscious awareness, of reflecting on that information and, where necessary, manipulating and modifying it. The buffer is episodic in the sense that it holds episodes whereby information is integrated across space and potentially extended across time. (...) it is assumed to be a temporary store (...) It is (...) assumed to play an important role in feeding information into and retrieving information from episodic long term memory (LTM).’ (p. 421). Figure 2.3 shows that the central executive and the episodic LTM may indeed interact. The latter is able to exchange information with the phonological loop and the visuo-spatial sketchpad via the language and visual semantics respectively.

As stated above, an important aspect of WM in (second) language listening is the phonological loop, since this is the component that deals with speech input. This component will therefore be described in detail in the following section.

**Phonological loop**

In every day life we often have to hold in mind non-visual and non-spatial information, like for instance a telephone number or a person’s name, for a short time. We try to facilitate this task by constructing a phonological representation of the input and by rehearsing this construct (out loud or in silence). The rehearsing process takes place in what is called the Phonological Loop. The Phonological Loop consists of two components:
(i) **a storage system**, i.e., the system for the representation of verbal information (in phonological or phonetic form) and

(ii) **a rehearsal mechanism.** The incoming material is encoded in the phonological store according to its sound-based characteristics. The auditory memory traces that are comprised in the temporary store decays within two seconds unless they are re-activated by the rehearsal mechanism.

Figure 2.4 shows a model of this Phonological Loop (Baddeley & Hitch, 1974, Baddeley 1986).

![Phonological Loop Model](image)

**Figure 2.4: Baddeley’s (1986) model of the phonological loop.**

In essence, the phonological loop is knowledge-free as phonological information is kept in the store, regardless of its correspondence to a familiar or an unfamiliar sound pattern. However, experimental work shows that the loop (and indeed the WM as a unit) is linked to language knowledge, and, as described in the previous subsection, the bridge between knowledge and the input is formed by the episodic buffer.

Newport’s ‘less-is-more’ hypothesis (1990) states that the benefit language learners have of a larger WM, forms, at the end, an obstacle for reaching high levels of achievement. This hypothesis may offer an explanation for the so-called critical period phenomenon, pertaining to an age-related decline in ability to acquire a foreign language (or even a first language). The argument Newport gives in favour of her hypothesis is that their large WM makes adult language learners focus on the input as a whole. As a result they do not make a detailed analysis of the encoded elements. The limited WM of children, however, makes it impossible for them to
process the input as a whole. They are forced to process the input in detail. According to the less-is-more hypothesis this eventually leads to a higher ultimate attainment, despite the fact that the acquisition process will initially be slower. Children’s detailed input analysis causes a better understanding of the internal structure of the language to be learned. However, as Miyake and Friedman (1998) state, the less-is-more hypothesis does not mean that a large WM is no longer important. The less-is-more principle compares the WM of young children with that of adults, while the theories discussed earlier in this section compare the WM capacities among adults. Research shows that some adult second-language learners are able to achieve a very high level of performance (for example in pronunciation), even if they started to learn the language at a later age (Bongaerts et al. 1997; Bialystok & Hakuta 1994). A larger WM might be the distinctive feature between the adults that do reach this near-native level and those that do not.

Experimental research indicates that the phonological loop has an important role in (i.e., highly correlates with) L1 vocabulary acquisition (Gathercole & Baddeley 1989) as well as in L2 vocabulary acquisition (Service 1992). Results of research with bilingual and monolingual children show that familiarity with a language influences the functioning of the phonological loop in a positive way (Gathercole & Thorn, 1998). Taken together, these studies indicate that the capacity of the loop seems to be dependent on the familiarity with the sound of the item and on (L2) language proficiency (Miyake & Friedman 1998). Therefore, the best way of testing the loop in order to establish its basic capacity is to use a test based on a language unfamiliar to the test taker. The absence of long-term representations of the items will make the outcome a more pure measure of the phonological loop.

2.5.2. Measuring working memory capacities

In the previous section it was said that both functions of WM, storage and analysis, play an important role in language processing. Since WM capacity is assumed to be limited, storage and processing have to compete with each other and trade-off effects might be expected (Daneman & Carpenter 1980). Since both storage and processing are important in WM it is important that

---

9 A study conducted by Geva & Ryan (1993) shows that the correlation between L2 WM and L2 proficiency is not just a reflection of general intelligence. The results of this study also suggest that the role of WM in L2 is stronger than in L1.

10 In our Training Study, we have chosen for a measure using unfamiliar words (nonwords) but familiar phonemes and phoneme combinations. See Chapter 5 for details.
both aspects are involved in measures of WM capacity. The two components that are always present in WM tests are:

(i) a measure to test whether the input information has been processed to a relevant degree, and

(ii) a measure of the information that has to be stored and remembered over the course of a test trial.

In most cases the so-called WM span is then operationalised as the maximum amount of information that the test taker can remember. In other words, the span is based on a measure of the second component. In this section several tests will be described that have been reported in the literature. The descriptions will be limited to tests of the WM as a whole, and tests of the phonological loop.\footnote{See Gathercole (1999) for tests to measure the visuo-spatial component.}

A test that is typically used as a measure of Working Memory is the Reading Span test, originally developed by Daneman and Carpenter (1980). In the test both components of working memory are involved: the processing component for dealing with the current sentence and the storage component for storing words to be remembered. In each test trial, participants are given some sentences that are presented one by one. Participants are asked to read them out loud and to try to remember the last word of each sentence. Reading involves the processing component of WM, while rehearsal of the sentence final words involves its storage component. Participants are told that the number of sentences per trial increases over the course of the test. Their reading span is defined as the maximum number of sentences that they can read while correctly recalling the last word. It is scored in a simple way by calculating the number of words correctly recalled.

There are several adapted versions of the reading span test. Differences between the derived and the original tests can be found in the modality of input presentation (e.g., written vs. oral) or in the sentence processing assignment (e.g., reading aloud vs. making a semantic judgment on sentence content). In the Training Study of this thesis a listening span test was constructed and used as a measure of working memory capacity differing from Daneman and Carpenter’s original reading span test in terms of modality. A detailed description of our listening span test is given in Chapter 5.

Examples of tests that are used for measuring the capacity of the phonological loop are the nonword repetition test, and the serial recognition test (Gathercole 1999; Gathercole et al. 2001). These tests rely on the storage
and rehearsal mechanism of the phonological loop and are based on the same principle as the reading span test:

(i) participants have to process the information to be able to give a correct answer, and

(ii) the amount of information that has to be processed increases over the course of the test.

In a digit span test, series of digits are presented which the participants have to repeat while in a nonword repetition test participants repeat series of nonwords. In a serial recognition test, a series of words or nonwords is presented twice. The order of the words (or nonwords) in the second presentation, however, may or may not remain the same as the order in the first presentation. Participants have to indicate whether they think the order is the same or different. An advantage of tests that makes use of nonwords is that they rely on LTM to a much lesser degree than when existing words are used. An additional advantage of the serial recognition test over a repetition test is that performance is not influenced by articulatory and phonological output abilities. Because of these assets a serial recognition test was used in the Training Study of this thesis. A detailed description of this serial recognition test is given in Chapter 5.

2.6. Automatisation

In this section the concept of automatisation is described. Section 2.6.1. gives some theoretical background and addresses problems that arise in defining the concept of automatisation. Section 2.6.2. presents the coefficient of variability, a possible gauge for measuring automatisation.

2.6.1. Theoretical background

Every time we try to learn a new skill (e.g., learning a new language, learning to play an instrument) we have to deal with the fact that, in the beginning, it is a time consuming and effortful process. However, we know that when we practice a lot, the process will become easier. After a certain period of (intensive) practice, the process runs off without effort, without awareness, and much faster than in the beginning. The process has then become automatised.

Notwithstanding an extensive literature on automaticity of cognitive processes, it is hard to find a clear definition of what automatisation exactly
is. One of the earliest definitions of an automatic process is given by Posner & Snyder (1975): ‘... the process occurs without intention, without giving rise to any conscious awareness, and without producing interference with other ongoing mental activity.’ (p. 56). Several characteristics have been proposed to distinguish between automatic and controlled processes. The typical criteria were: the speed of the processes, the error rate, the effort required, the capacity used, and the level of possible control. Later, researchers rejected several of these criteria as unnecessary or insufficient to make a distinction between the two kinds of processes. The last twenty years have witnessed several developments in automatisation research. First, the idea emerged that controlled and automatised processes are not mutually exclusive. Most researchers now posit a continuum of automatic processes at one extreme, controlled processes at another, and many processes of a mixed nature in between. Second, a shift occurred from theories that reflect automatisation in relation with how much attention was required to theories that present automatisation as a reflection of how memory is used. Third, views changed with regard to when a process has become automatised. Some researchers ascribe automatisation to the simple speed-up of component processes (Carlson et al. 1989; Logan 1988) while others define automatisation not as a simple speed-up (a quantitative change) but as a qualitative change in the composition of component process (Segalowitz & Segalowitz 1993). The emphasis that Segalowitz lays on the qualitative change makes his position on automatisation rather unique (Segalowitz & Segalowitz 1993; Segalowitz & Gatbonton 1995; Segalowitz, Segalowitz & Wood 1998; Segalowitz 2000). Segalowitz, nevertheless, subscribes to the practical importance of improved speed and accuracy for achieving fluency in language use. He makes a distinction between ‘performance fluency’, pertaining to the observable speed and accuracy of component processes, and ‘cognitive fluency’, pertaining to the efficiency with which various component processes are blended and balanced (Segalowitz et al 1998; Segalowitz 2000). After intensive practice the blend will change in favour of automatised processes. Performance will then become faster and more stable because the influence of or the reliance on the relatively slow and variable controlled processes is eliminated or reduced. However, when necessary, an otherwise automatic process, such as the pronunciation of planned utterances

12 There are some positions that lie between the two approaches mentioned here. Cheng (1985) for example describes automatisation as a reorganization of components, whereas both Shallice (1982) and Stanovich (1990) characterize automatisation as a shift from reliance on higher-order to a more independent processing mechanism, taking place in what Fodor (1983) calls ‘informationally encapsulated’ processing units. In such units, information processing finds place without involvement of higher- or lower-level information.
in a language that we command fluently, can be held under conscious control, e.g., when we want to pronounce certain words or sounds in an uncommon way. Or, in the case of receptive language use, we can, if the situation demands, pay special attention to the way someone else pronounces certain words or sounds, which would otherwise go by unnoticed.

To determine the place of automatisation in skill development, two theories are often cited in the literature on second language acquisition, namely the rule-based ACT-R theory of Anderson (1993) and the item-based Instance theory of Logan (1988).\textsuperscript{13} Important concepts of the ACT-R theory are declarative knowledge (e.g., the knowledge that ‘work’ is a verb and that ‘Petra’ is a proper noun) and procedural knowledge. Procedural knowledge has the form of production rules of the \textit{if x then do y} type (e.g., if the subject of a sentence is taken by a proper noun with the features third person and singular, and if the predicate has the feature present tense, the verb takes the suffix -s, leading to the computation of sentences like ‘Petra works’). The idea is that in early skill acquisition, performances are based on production rules to retrieve the relevant knowledge out of declarative memory and putting them together in working memory. Due to practice, components of the skill that are repeated can become routinised, forming chunks. These chunks can be retrieved fast and efficiently; they are used relatively independently of conscious control. Thus, procedural knowledge can be said to become increasingly automatic.

In Logan’s (1988) Instance theory, automatisation means a shift from algorithm-based performance to memory-based performance. In this theory ‘\textit{performance is automatic when it is based on single-step direct-access retrieval of past solutions from memory}’ (p. 493). Each time the algorithm is used, a new memory trace corresponding to the rule is formed. The more the rule is used, the more memory traces there will be. Each time an action has to be performed, there will be a race between the original algorithm and the memory traces. This race will eventually be won by direct access to memory traces. The raised activation levels of the memory traces, being the result of increased experience, will make it more efficient to retrieve the traces than to execute the algorithm itself.

\subsection*{2.6.2. Measuring automatisation}

How now to measure automatisation? All theories of automatisation regard speed of processing as an important indicator of automatisation. This can be

\textsuperscript{13} The ACT-R theory was preceded by the ACT* theory (1983) that was preceded by the original ACTE theory (Anderson, 1976). ACT stands for Adaptive Control of Thought.
reflected in the speed with which a behavioural response is given to a stimulus. However, Segalowitz and Segalowitz (1993) claim, as stated above, that simple speed-up is not a true sign of automatisation since it does not necessarily involve a restructuring of sub-processes. In their 1993 paper they propose the ‘coefficient of variability’ (CV<sub>RT</sub>) as a measure for automatisation. This ‘coefficient of variability’ is defined as the standard deviation (SD) of reaction times divided by the mean reaction time (RT) of responses in, for example, a lexical decision task. It is a measure of variability that is corrected for the latency of responding. In case of a mere speed-up, the standard deviation (SD<sub>RT</sub>) is reduced proportionally to the reduction of the RT. Thus, there will be no change in the relative variability of RT (CV<sub>RT</sub>). In the case of restructuring, however, some time-consuming controlled processes will be eliminated or modularised. For example, skilled readers do not pay attention anymore to the form of letters; they can quickly recognize words like bad, dab, dad, and baby, without paying attention to the difference in form of the letters 'b' en 'd'. Children learning to read, however, have to invest considerable time and effort to master the automatic recognition of these letters, paying special attention to the components of their forms. Skilled readers will omit these initial component processes. Consequently, not only will the overall RT of word recognition be reduced but also so will the SD<sub>RT</sub> be reduced (more than proportionally). As a result, the CV<sub>RT</sub> will reduce as well. The reduction of the CV<sub>RT</sub> can be regarded as a first criterion for determining automatisation. In addition to this CV<sub>RT</sub> reduction, another criterion has to be met before the conclusion that a process became automatised can be drawn. There has to exist a positive correlation between RT and CV<sub>RT</sub>. Because in the case of restructuring (automatisation) every reduction in RT will be accompanied by a more than proportional reduction in SD<sub>RT</sub>, CV<sub>RT</sub> should be positively correlated with overall RT. This means that a reduced CV<sub>RT</sub> combined with a positive correlation between CV<sub>RT</sub> and RT can be seen as an index of automatisation.

Let us take as an example a spoken word recognition task that is performed by a number of individuals a number of times in a period of several days. If automatisation were solely a matter of speed-up, then we would expect a shortening of reaction times during the course of the experiment. If, however, automatisation were a matter of quantitative as well as qualitative change, we would expect that certain component processes,

14 Before Segalowitz started to use CV<sub>RT</sub>, this index was not generally used in cognitive psychology research. It had, however, been used in other areas such as the study of action patterns in animal behaviour.

15 See Segalowitz et al. 1995 for a description of using the CV<sub>RT</sub> in a single-case study.
such as the recognition of the sounds and phonemes, would be reduced, compiled or even omitted altogether. This would result in faster reaction times but also in a more than proportionally reduced SD. As a result there would be a decline of the $CV_{RT}$. In this thesis we will use the $CV_{RT}$ to make a distinction between simple speed-up and automatisation (see Chapter 5).