Determinants of success in native and non-native listening comprehension: an individual differences approach

Andringa, S.; Olsthoorn, N.; van Beuningen, C.; Schoonen, R.; Hulstijn, J.

DOI
10.1111/j.1467-9922.2012.00706.x

Publication date
2012

Document Version
Final published version

Published in
Language Learning

LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

Determinants of Success in Native and Non-Native Listening Comprehension: An Individual Differences Approach

Sible Andringa
University of Amsterdam

Nomi Olsthoorn
University of Amsterdam

Catherine van Beuningen
University of Amsterdam

Rob Schoonen
University of Amsterdam

Jan Hulstijn
University of Amsterdam

The goal of this study was to explain individual differences in both native and non-native listening comprehension; 121 native and 113 non-native speakers of Dutch were tested on various linguistic and nonlinguistic cognitive skills thought to underlie listening comprehension. Structural equation modeling was used to identify the predictors of individual differences in listening comprehension and to test for differences between the native and non-native participants. Listening comprehension for native speakers was

---

Nomi Olsthoorn is now at the Department of Psychology, Lancaster University, Lancaster.

This research was funded by the Netherlands Organisation for Scientific Research by a grant awarded to Hulstijn (NWO grant 360–70-230). We thank our research assistants Tineke van der Linde and Netta Meijer for their splendid help in collecting and organizing the data, and Dirk-Jan Vet for his technical assistance. We would also like to thank Antje Meyer, Leah Roberts, and two anonymous Language Learning reviewers for their careful reading of our text and their valuable suggestions for improvement.

Correspondence concerning this article should be addressed to Sible Andringa, Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 134, 1012 VB, Amsterdam, the Netherlands. Internet: S.J.Andringa@uva.nl

found to be a function of knowledge of the language and the efficiency with which one can process linguistic information, while listening comprehension for non-native speakers was a function of knowledge and reasoning ability. Working Memory did not explain unique variance in listening comprehension in either group. Differences in experience with the Dutch language are likely to explain the observed pattern of results for both groups.

Introduction

What does it mean to be a proficient listener? Upon hearing speech, listeners tap into all kinds of linguistic knowledge and skills in order to make sense of the incoming message. They isolate speech from background noise, segment the speech stream and perceive word boundaries, select words from an array of activated candidates, and use the syntactic and semantic information carried by individual words to integrate them into the larger sentence context. In addition, sentence meanings are integrated into the ongoing discourse (Cutler & Clifton, 1999). As listening is an online activity, it is essential that these processes run fast and efficiently. Apart from linguistic abilities, general cognitive skills—most notably the ability to store and process linguistic information—are also thought to be involved in the listening process. Being a good listener requires all of these resources and skills. Thus, individual variation in listening comprehension success, which has been shown to exist in the large body of research on the role of working memory in comprehension (for a review, see Daneman & Merikle, 1996), may be due to differences in linguistic knowledge, differences in people's capacities to efficiently engage in the complex set of processes underlying listening, or differences in general cognitive ability. The goal of the present study was to identify and compare these sources of individual differences in the listening skills of both native and non-native speakers. By identifying the factors that explain individual differences in listening, we not only aimed to increase our understanding of why some people are more successful comprehenders than others, but we also strived to gain insight into the actual construct of listening comprehension.

So far, most listening research has adopted an experimental methodology in which aspects of listening are studied in isolation and variation amongst participants is reduced as much as possible. This exploratory study adopted a differential or individual differences approach. This should not be taken to mean that we investigated the effect of motivation, aptitude, learning style, etc., which are factors that are strongly associated with the term *individual differences* in the second language acquisition (SLA) literature. This was an individual differences study in the classical sense (Cronbach, 1957). Such a differential approach, where variation between individuals is of key interest and which implies a correlational research design, has rarely been employed to study listening, apart from a substantial body of work investigating the role of working memory in listening (and reading) comprehension (see also Roberts, 2012).

In the remainder of the introduction, we will review (1) insights from experimental sentence processing studies comparing native and non-native listeners, (2) the few investigations into listening comprehension that have adopted a differential approach, and (3) research on the relationship between working memory and comprehension success.

Studies Comparing Native and Non-Native Listening Comprehension Processes

The processes underlying listening comprehension are served by many types of linguistic (and non-linguistic) knowledge. Many studies of speech processing have been devoted to uncovering the cues that listeners use to successfully comprehend speech (e.g., Altmann & Steedman, 1988; Cutler & Norris, 1988; Federmeier, 2007; McClelland, 1986; Norris, McQueen, Cutler, & Butterfield, 1997; Trueswell, Tanenhaus, & Garnsey, 1994). This research has shown that successful comprehension is the result of both bottom-up and top-down processes, that is, it is driven both by observable cues in the input, by cues implied by the selected lexical candidates, and by expectations based on the listener's general (world) knowledge and/or the previous context.

Studies comparing the exploitation of linguistic cues in native and non-native listening comprehension have shown that the comprehension processes of the two groups differ in two important respects. First, there is evidence that native and non-native listeners do not rely on the same cues in speech processing: non-native listeners appear to be less able to make use of syntactic information. For example, Sanders, Neville, and Woldorff (2002) presented native and advanced non-native speakers of English with sentences that were reduced in terms of the amount of lexico-semantic and/or syntactic information they carried, and found that the non-native listeners were not able to use syntactic information as a cue in a phoneme monitoring task to the same extent as native speakers. Similarly, in a reading-time study, Papadopoulou and Clahsen (2003) observed that their advanced Greek learners of English relied more on lexical than syntactic cues in comparison to the native speakers. Based on such findings in both listening and reading, Clahsen and Felser (2006) proposed the shallow structure hypothesis, according to which non-native listeners cannot exploit syntactic information as adequately as native listeners can in comprehension, but construct less detailed syntactic representations (see also Roberts, 2012). In the field of SLA, several studies have investigated how learners deal with gaps in their linguistic knowledge. These studies suggest that L2 learners rely more on top-down cues in listening than native speakers (Field, 2004; Koster, 1987; Mack, 1988; Wolff, 1987).

Second, non-native speakers may differ from native speakers in the efficiency or degree of automaticity with which they use linguistic cues. For instance, in investigating the ability of Dutch learners of English to process prosodic information, Akker and Cutler (2003) found that learners were able to do this adequately—the non-native listeners were slower, but their response latencies fell within the native speaker range. However, they also demonstrated a different response pattern to targets in prosodically accented or unaccented sentences, even though English and Dutch do not differ in this respect. Akker and Cutler speculated that this was due to non-native listeners being less efficient in relating prosodic to semantic information. Van Hell and Tokowicz (2010) put forward a similar suggestion when they pointed to the fact that most Event-related potential (ERP) studies of second language (L2) grammatical processing have not found an ELAN (early left anterior negativity) effect for non-native listeners, which is thought to reflect highly automatic structure building processes.

This review demonstrates that differences in linguistic knowledge and the efficiency in the application of such knowledge may explain some of the differences in the success of comprehension observed between native and non-native listeners. The question is whether these differences also explain variation *within* groups of native or non-native listeners.

Studies Into the Componential Structure of Listening Comprehension

Studies of the componential nature of a proficiency construct have been undertaken quite often for reading, writing, and speaking (e.g., De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012; Jincho, Namiki, & Mazuka, 2008; Schoonen, Hulstijn, & Bossers, 1998; Schoonen et al., 2003; Schoonen, Van Gelderen, Stoel, Hulstijn, & De Glopper, 2011; Van Gelderen et al., 2004). These studies tend to find that linguistic knowledge components as well as speed or fluency factors are associated with success in both first language (L1) and L2 reading (Van Gelderen et al., 2004), L1 and L2 writing (Schoonen et al., 2003), and L2 speaking (De Jong et al., 2012). Metacognitive knowledge or strategies were also found to explain success in reading (Van Gelderen et al., 2004) and writing (Schoonen et al., 2003). Jincho et al. (2008) included a reading span task in their study of reading comprehension, and found that this measure predicted variance in reading beyond what was explained by vocabulary knowledge and orthographic knowledge.

The differential approach has been far less common in listening comprehension research; to the best of our knowledge, there are only two studies—both from the field of SLA—that aimed at identifying components in listening comprehension. The first is a study by Mecarrty (2000) that sought to assess the (relative) contribution of lexical and grammatical knowledge to the listening proficiency of 77 L2 learners of Spanish. Whereas both types of knowledge were found to be significantly correlated to listening, only lexical knowledge proved to explain unique variance (i.e., 14%) in listening comprehension. Second, Vandergrift, Goh, Mareschal, and Tafaghodtari (2006) report a significant relationship between metacognition and L2 listening comprehension success of 115 L2 learners of English and 226 L2 learners of French; they found that self-reported metacognitive awareness and strategic competence accounted for 13% of the variability in listening comprehension.

It is clear that explaining individual differences in listening has not been high on the agenda, except for the large body of work investigating the role of working memory in comprehension.

Studies on the Role of Working Memory in Listening Comprehension

Most theories of comprehension assume that comprehension depends on the ability to temporarily store and perhaps process information in memory (e.g., Clark & Clark, 1977; Ericsson & Kintsch, 1995; Kintsch & Van Dijk, 1978; Kintsch, 1998; Townsend & Bever, 2001). Daneman and Carpenter (1980) were the first to provide support for this claim. They developed reading and listening span tasks where participants read or listen to series of unconnected sentences (which incrementally increase in number), make plausibility judgments, and have to remember the last word of each sentence in the series. The maximum number of final words correctly recalled determines the size of their reading or listening span. Daneman and Carpenter found substantial correlations between these tests and native speakers' performance on discourse comprehension tasks. These findings have been replicated often (for a review, see Daneman & Merikle, 1996). In addition, some studies have found similar correlations for non-native speakers (e.g., Harrington & Sawyer, 1992; Payne, Kalibatseva, & Jungers, 2009).

Exactly how working memory constrains comprehension is subject of debate. Working memory is often considered to be a single resource upon which different linguistic processes draw simultaneously, and, as a result, comprehension is constrained by the amount of capacity available in the working memory system (Just & Carpenter, 1992; King & Just, 1991; MacDonald, Just, & Carpenter, 1992). Waters and Caplan rejected this view, claiming that a specifically dedicated memory system handled syntactic parsing (Caplan & Waters, 1999; Waters & Caplan, 1996). In their view, reading span tasks index consciously controlled processes rather than the implicit and automatic processes at work in sentence comprehension. They interpret correlations such as those reported by Daneman and Carpenter (1980) as the result of post-interpretative (post-parsing) processes. MacDonald and Christiansen (2002) also claimed that there is no role in comprehension for a working memory system that is functionally separate from the cognitive operations executed. They look at language processing from a connectionist perspective and claim that working memory does not exist as such. In their view, individual differences in comprehension are mostly the result of experience: "some individuals engage in language comprehension, particularly reading, more than others" (p. 36). Reading and listening span tasks are not measures of verbal working memory capacity; they are measures of the ability to engage the language processing system, which is affected by experience. Thus, working memory and language processing skill are each the flip-side of the same coin.

To tease apart the single resource and the specifically dedicated working memory accounts, researchers have tried to demonstrate that parsing is or is not affected by verbal working memory. In the dedicated working memory view, parsing should not be affected by verbal working memory as measured by span tasks, whereas it would be in the single resource account. As matters stand, it is probably too soon to tell which account is correct. Much of the evidence available so far is based on reading research and on one syntactic phenomenon – the processing subject and object relative clauses (*it was the scream that frightened the boy* versus *it was the boy that the scream frightened)*, which are assumed to be differentially costly in terms of parsing and consequently also in the extent to which they draw on working memory capacity. Also, the evidence is mixed. Some researchers report that there is no relationship between the ability to parse and working memory capacity (e.g., Waters, Caplan, & Yampolski, 2003; Waters & Caplan, 2004), but most find that working memory capacity moderates the processing of sentences that are somehow ambiguous (e.g., Fallon, Peelle, & Wingfield, 2006; Havik, Roberts, Van Hout, Schreuder, & Haverkort, 2009; MacDonald et al., 1992; Pearlmutter & MacDonald, 1995; Traxler, Williams, Blozis, & Morris, 2005). Effects of working memory additionally appear quite consistently when the comprehension task taps higher order interpretation processes (e.g., Daneman & Carpenter, 1983; Daneman & Green, 1986; Masson & Miller, 1983).

There is also evidence for the experience-approach to working memory. To investigate whether the cost associated with sentence processing is a function of one's experience, Reali and Christiansen (2007) also studied the processing of subject and object relative clauses. They first investigated the distributional properties of these clauses by means of corpus research, and they found that object relative clauses are more likely to occur when the embedded noun phrase is a personal pronoun, whereas subject relative clauses are more frequent when the embedded noun phrase is impersonal. They subsequently explored how participants responded to violations of these patterns in a series of self-paced reading experiments. They found that clause type (subject or object relative clause) is not the determining factor: processing cost was a function of whether or not the sentences were in line with the observed distributions in the corpus. Wells, Christiansen, Race, Acheson, and MacDonald (2009) claimed that the commonly observed differences in processing cost for subject and object relative clauses are due to the fact that people have not been exposed much to the kinds of object relative clauses that are commonly used in most studies, simply because these do not occur much in natural language. They developed a training experiment in which they exposed participants to both subject and object relative clauses, and expected that the instruction would affect the processing object relative clauses only. This expectation was confirmed, and they concluded that individual differences in sentence processing may be due to previous processing experiences.

Rationale for the Present Study

As reviewed above, individual differences in listening comprehension have not yet been studied extensively. On the basis of group comparisons of native and non-native listeners, it seems that differences in linguistic knowledge and/or the efficiency with which such knowledge can be applied may explain differences in listening success. In addition, while working memory capacity is a known predictor of native listening comprehension ability, the exact nature of the relationship is still unclear. The question of whether and if so, the extent to which working memory explains non-native comprehension is also unresolved. In addition, to our knowledge, the role of working memory in listening comprehension has not been studied in relation to other predictors, such as linguistic knowledge and speed of processing. The present study adopts an individual-differences approach to further our understanding of listening comprehension by conjointly assessing the importance of linguistic knowledge, processing speed of linguistic information, and general cognitive ability for both native and non-native speakers. The overall research question was:

RQ: Which factors explain success in native and non-native listening comprehension, respectively?

This research question subsumes several more specific questions, such as: What is the relative role of (1) linguistic knowledge, such as knowledge of vocabulary; (2) the efficiency with which this knowledge can be processed; (3) cognitive abilities (working memory and reasoning skills) in listening comprehension; and (4) to what extent are the observed patterns different for native and non-native speakers?

Method

Design

The aim of this study was to explain individual differences in native and non-native listening comprehension. Our approach was to operationalize potential sources of individual differences—measures of the knowledge underlying comprehension processes and the speed with which such knowledge can be applied, and measures of general cognitive ability—and to assess their contribution to the explanation of individual differences in native and non-native discourse comprehension. Native and non-native speakers of Dutch completed a large number of tasks assessing their Dutch-language skills, working memory, and verbal reasoning ability. The constituent variables were measures of (1) the linguistic knowledge that is required to construct sentence meaning, (2) efficiency in applying that knowledge (processing speed), and (3) cognitive ability as indicated by working memory capacity and reasoning ability (IQ). All measures were the same for native and non-native speakers. Multisample structural equation modeling (SEM) analyses were performed to answer our research questions.

Participants

The participants in this study were 121 native speakers of Dutch and 113 speakers of Dutch as a second language. Studies of individual differences presuppose the existence of notable individual differences (i.e., sufficient variance within the groups). To ensure heterogeneity in language proficiency, participants of both high (higher vocational and university) and low (all other vocational) educational backgrounds were recruited (for the native speakers 61 high and 60 low; for the non-native speakers, 66 high and 47 low). The age range was the same for both groups, from 19 to 40, but the native listeners were slightly younger (*M* = 25; *SD* = 5.1) in comparison to the non-native listeners (*M* = 29; *SD* = 5.2). In both groups, women were overrepresented (84 over 37 in the native group and 76 over 37 in the non-native group). The L2 learners varied in Dutch proficiency from B1 on the Common European Framework of Reference for languages (CEFR) (intermediate user) to near-native levels. Their mean age at arrival in the Netherlands ranged from 0 to 36 (*M* = 21; *SD* = 8.1), and their length of stay in the Netherlands varied from 8 months to 27 years (*M* = 8; *SD* = 6.1). Thirty-five different first languages were present in the non-native speaker sample; the most frequently occurring languages were German (9 speakers), Russian (9 speakers), Bahasa Indonesian (9 speakers), and Spanish (8 speakers). Participants received 40 euros for participation.

Procedures and Tasks

It took participants four hours to complete all tasks. Two sessions of two hours were organized, and with a few exceptions, these sessions were held on separate days. The tasks were administered in a fixed order. At the start of the first session, participants were informed about the purpose of the study, filled out background questionnaires, and signed consent forms.

Before each task, participants received oral and written task instructions, they were given the opportunity to practice on several trials, and they had the opportunity to ask questions. Most tasks were computerized. Participants were seated at about 60 cm from a computer screen while wearing headphones. Special programs were developed to administer the discourse comprehension task and the vocabulary task. The semantic processing, grammatical processing, segmentation, word monitoring, and self-paced listening tasks (see below for details) were set up and run with the E-prime software package (Schneider, Eschman, & Zuccolotto, 2002), and their responses were logged by means of a keyboard. The utterances presented in these tasks were read at a normal pace by a female native speaker of Dutch with a neutral accent and recorded in a sound proof studio at 16 bit, 44 Khz. Care was taken to retain the characteristics of natural speech: the sentences in these tasks were mostly inspired by—and sometimes literally taken from—a corpus of spoken Dutch: *Het Corpus Gesproken Nederlands (2006)*.

Discourse Comprehension

The dependent variable in this study was a discourse comprehension test that consisted of five texts taken from the Dutch State Exam of Listening Proficiency, a national exam for advanced learners of Dutch as a second language.

Participants had to listen to five different conversations, between male and female speakers, which took the form of an interlocutor asking short questions that were answered in short monologues by an expert on a particular topic. Thus, each conversation consisted of a number of question-and-answer pairs, and one multiple choice (MC) question was asked about each. The MC questions were presented on paper and each question had three response alternatives. Following the state exam procedures, the test was timed. After each question-and-answer fragment, participants had at most forty seconds to answer the MC question and to prepare for the next. Participants were explicitly instructed to do so. There was no opportunity to pause or replay the text fragments. The test consisted of 36 items which were presented in a fixed order and administration took one hour and ten minutes. The number of correct responses were scored. Participants were given a short break between the second and third text.

Vocabulary

To assess vocabulary size, a receptive computer administered multiple choice test was used, based on a selection of items from the Hazenberg and Hulstijn (1996) vocabulary test. As this test was originally intended for non-native speakers only, additional and more difficult items were constructed following the principles of the original test to make it suited for assessing native speaker vocabulary as well. The test consisted of a selection of 60 items from a total of 140 that were piloted amongst 38 native and 13 non-native participants. The items were selected both on the basis of frequency information from the Celex corpus (Baayen, Piepenbrock, & Gulikers, 1995), a large corpus of written Dutch, as well as their psychometric properties as derived from the pilot study. In the test, each target word was presented in a simple carrier sentence that did not reveal its meaning. Care was taken not to introduce any systematicity in the length of the alternatives and the way meanings were described. Participants had five alternatives to choose from, the last one always being: “I really don’t know.” The other four alternatives consisted of one correct and three incorrect definitions of the word meaning. They were constructed in simple, high frequency language. An accuracy score was calculated based on the total number of correct responses. Administration took approximately 20 minutes, but participants were allowed to take more time.

Semantic Processing

This task was intended as a measure of the ability to comprehend single utterances. Participants were required to listen to single utterances, and demonstrate comprehension by choosing the communicatively adequate response from two alternatives. These responses always consisted of highly frequent words or phrases and were maximally four words long, such as "thank you" or "I don't know." The two alternatives were presented simultaneously on the left and the right of the screen for 3 seconds before and while the utterance was played. For example, after seeing the response alternatives *Nee sorry* ("No, sorry") and *Goed idee* ("Good idea"), participants heard: *Gaat de bus naar Amsterdam nog wel?* ("Is there still a bus to Amsterdam?"). Participants were required to press the appropriate button depending on which alternative they considered appropriate. The task consisted of four practice and 56 test items. Sentences varied in lexical frequency of the words used (containing only highly frequent words vs. containing at least three words of low frequency), syntactic complexity (simple vs. complex) and length (short vs. long). Both response accuracy and latency (from sentence onset) were logged. Administration took approximately 10 minutes. Accuracy scores for this task were not used because participants performed at ceiling level.

Grammatical Processing

This task tested participants' knowledge of the distributional and combinatorial properties of the Dutch language. In the task, participants heard sentence beginnings and were instructed to indicate as quickly as possible whether the fragments were possible sentence-initial strings in Dutch by pressing the correct (Yes) or incorrect (No) button. The sentence beginnings were kept short (three to four words) to avoid taxing memory. For example, hearing *Die stad lijkt heel* ("That city seems very") should evoke a yes-response, whereas hearing *Precies ik weet* ("Exactly I know") should evoke a no-response. The task consisted of six practice and 34 test items. The sentence beginnings were three to five words long. Half of them were correct; the other half consisted of strings that are not permissible (in sentence initial position) in Dutch because of word order or agreement violations. Both response accuracy and latency (from fragment onset) were logged. Administration took approximately 5 minutes.

Segmentation

This task was intended as a measure of segmentation ability, the ability to recognize words in the speech stream. Participants heard short fragments and were instructed to respond as quickly as possible by choosing a number from 1 to 5 corresponding to the number of words they thought they had heard. To limit the influence of working memory, strings consisted of 2 to 4 words; there were no strings of one or five words. The task consisted of 60 items (3 practice and 57 target items). Half of them were fully articulated, and the articulation of the other half was reduced according to the principles of vowel reduction or consonant reduction (or both) as described in Ernestus (2000), Kloots, De Schutter, Gillis, and Swerts (2003), and Coussé, Gillis, and Kloots (2007), typical of normal colloquial Dutch (compare *ik heb het* and *kepət,* meaning “I have it”). The reductions used in the items were actually observed in *Het Corpus Gesproken Nederlands* (2006). Fragments always consisted of high-frequency words only: If the task was difficult, it was due to the reductions. Administration took approximately 10 minutes.

Word Monitoring

This task tested participants’ ability to use their knowledge of the distributional and combinatorial properties of the Dutch language to predict upcoming information. In each trial, a target word to be monitored was printed on screen for 1 second and participants were instructed to remember that word. Subsequently, a carrier sentence was played and participants were instructed to press the space bar as soon as they heard the target word. The monitor task consisted of 46 experimental items and 4 practice items. The sentences varied in length from 7 to 17 words, and sometimes consisted of main clauses only, and sometimes of main clauses with coordinate, subordinate, or relative clauses. The position of the word to be monitored in the carrier sentence varied from second to fourteenth position. The words to be monitored were nouns (23), verbs (8), adjectives (6), prepositions (5), and adverbs (4), and varied in degree of predictability. All carrier sentences consisted of simple, high frequency words. Response latency was logged from the onset of the word to be monitored. Participants could press before the word actually appeared; as a result, negative latencies were possible. Administration took approximately 5 minutes.

Self-Paced Listening

The word-by-word self-paced listening task was included as a measure of sentence processing efficiency, the assumption being that fast pacing reflects efficient comprehension. The task was set up in accordance with the specifications provided by Ferreira, Henderson, Anes, Weeks, and McFarlane (1996). Participants listened to sentences one word at a time (articles were presented with the next word) at their own pace at the press of the space bar. If the space bar was pressed before the end of a word, it would be truncated and the next word would be played. Participants were instructed to pace as fast as they could without losing track of what was said. One in three sentences was followed by a simple yes-no comprehension question to promote faithful execution of the task. The self-paced listening task consisted of 56 items and 4 practice items. The items varied in lexical frequency, syntactic complexity, and length (as in the semantic processing task above). Although sentences were presented and recorded word by word, care was taken to retain Dutch intonation contours as much as possible. Response latencies per trial were determined by averaging the response latencies per word from word offset. Responses to sentence-initial and sentence-final words were excluded from the analyses. Administration took approximately 15 minutes.

Working Memory

The reading and listening span tasks as proposed by Daneman and Carpenter (1980) have been claimed to be measures of proficiency or experience more than working memory capacity (MacDonald & Christiansen, 2002). For this reason, four digit span tasks (forward and backward visual and forward and backward auditory) and one non-word recognition task were used as measures of verbal working memory capacity. For the digit span tasks, participants were asked to look at or listen to a series of digits that appeared one by one, and to reproduce the series by typing in the digits they saw or heard in the same (forward) or reverse (backward) order. The digits were presented in 75-point Arial font in blue on the centre of a white computer screen for the duration of 1 second each, after which the next number was presented immediately. For the auditory versions, participants listened to Dutch digits, artificially generated by means of Fluency text-to-speech software (Dirksen, 2008) and edited in PRAAT (Boersma & Weenink, 2005) to be each exactly one second in duration. The minimum series length was two digits, increasing with one digit every two trials until the maximum length of eight (backward) or nine (forward) digits was reached, or until participants failed to respond correctly on both trials of a particular length. The tasks were administered separately, in between the other tasks, and took about 5 minutes each to make.

In the nonword recognition task, participants listened to series of consonant-vowel-consonant (CVC) pseudowords of increasing length. Each item in the task consisted of two series of pseudowords, and participants had to decide whether the two series they listened to were the same or different. The length of the series increased as participants progressed in the task. The task comprised 21 trials, three of each length, starting at length two and progressing to length eight. Every trial played back the series of pseudowords at a rate of one per second, with a 400 ms pause between items. The last item of a series was followed by a 1500 ms pause, after which the repetition series was played back. The last item of the repetition series was followed by a screen prompting the question whether the two series were identical, which participants had to answer by pressing the correct (Yes) or incorrect (No) button. The repetition series always consisted of the same pseudowords as the original series, either in the same order (requiring a "yes" response) or with the order of two adjacent items reversed (requiring a "no" response). Participants did not receive feedback on their responses. The total duration of the task was about 10 minutes.

Intelligence

Intelligence was assessed with the complex matrices component of the Wechsler Adult Intelligence Scale, third edition, a measure of nonverbal IQ (Wechsler, 1997). The test was included to assess reasoning ability, and was administered in accordance with WAIS-III specifications. The items were presented on paper and participants indicated their solution on an answer sheet.

Analyses

The first step in the analyses was to inspect each measure for outliers. Responses were considered outliers when they fell outside the range of 2.5 standard deviations from the mean, which was calculated per item for both groups separately. Outliers were set to missing. This never constituted more than 2.4% of the data points in any task. In addition, for semantic processing speed, grammatical processing speed, and segmentation speed, latencies to inaccurate responses were considered invalid and also set to missing (2.4%, 9.5%, and 18%, respectively). Some individuals made many mistakes. If more than 40% of a participant's responses were incorrect, it was assumed that the task had not been understood or that the participant's proficiency was too low and the score was removed all together (one non-native speaker for semantic processing speed, five non-native speakers for grammatical processing speed, and one native speaker and four non-native speakers for segmentation speed). All missing values were then imputed in SPSS by means of the full information maximum likelihood estimation procedure, and mean scores per subject were calculated.

Multisample SEM was used to analyze the data. SEM combines confirmatory factor analysis (CFA) and regression. CFA is used in SEM to determine whether the measured scores, the observed variables, group together under predetermined latent variables or factors. These factors should reflect theoretically motivated constructs. In our study, Listening Comprehension, Knowledge, Processing Speed, and Cognitive Ability were the factors assumed to underlie the measures presented above. SEM provides measurement regression weights for each observed variable, which are indications of the strength of the association between the observed variable and the latent factor, and it tests whether these weights are significantly different from zero. SEM also uses correlation and regression techniques to model how the extracted factors are related to each other. In this study, Listening Comprehension was regressed onto Knowledge, Processing Speed, and Cognitive Ability factors as a reflection of the idea that the latter three factors are possible sources of individual differences in listening comprehension. The strength of the relationship between latent variables is indicated by structural regression weights and accompanying significance tests. SEM also allows for hypothesized models to be fitted to more than one group, and it tests for differences in regression weights between the groups, which is how the native and non-native listeners were compared in this study. Differences would indicate that particular relationships interact with group membership (Kline, 2005).

An advantage of using SEM is that measurement error in the observed variables is reduced. The factors that result from the CFA consist of the variance that each of its indicators (the observed variables) have in common. Thus, measurement error in the observed variables is partialed out and the latent variables may be considered free of measurement error. Because latent variables are preferably indicated by at least three observed variables, the items of all tasks in this study were split randomly into three subsets or parcels (except for the working memory tasks, for which we had five separate measures already). For example, the segmentation items were divided over subsets A, B, and C. A segmentation accuracy factor was made up of the variance in accuracy scores that A, B, and C had in common. Because task items were split into subsets, reliability coefficients under a congeneric model were calculated instead of Cronbach's alpha according to the procedures described in Fleishman and Benson (1987).

To answer our research questions, the data were analyzed according to the steps recommended for multisample SEM by Byrne (2010). First, descriptive statistics were obtained to inspect the data and assess normality. First-order CFAs were performed to assess whether our tasks measured separate constructs and to assess their reliabilities. Then, SEM procedures were used to assess which configuration of factors fitted the data, and whether the same configuration would hold for both groups. Model fit was tested statistically by a χ^2-test. However, in large samples and complex models this test is not very informative, and descriptive measures are usually preferred, such as the χ^2/df ratio (preferably < 2), the comparative fit index (CFI; preferably $> .95$), and the root mean squared error of approximation (RMSEA, preferably $< .06$) (cf. Hu & Bentler, 1999). Tests for differences between the regression weights for native and non-native speakers were conducted by testing for group invariance. This was done by setting relationships in the model to be equal for both groups. The validity of such equality constraints was assessed by determining whether the loss in chi-square points was significant compared to the loss of degrees of freedom.

Table 1 Means, standard deviations, and partial η^2 for the differences in native and non-native mean scores

Measure	Native Mean (s)	Non-native Mean (s)	Partial η^2
Dependent variable			
Listening comprehension (Max = 36)	33.6(2.0)	28.3(6.9)	.22
Knowledge measures			
Vocabulary (Max = 60)	40.7(6.3)	28.1(9.4)	.39
Grammatical proc. acc. (Max = 34)	32.4(1.4)	28.9(4.4)	.23
Segmentation accuracy (Max = 60)	39.2(3.8)	36.8(5.9)	.06
Processing speed measures			
Semantic processing speed (ms)	3493 (606)	4346 (838)	.26
Grammatical processing speed (ms)	1488 (242)	2020 (570)	.28
Segmentation speed (ms)	1338 (408)	1975 (718)	.23
Word monitoring (ms)	278 (59)	370 (88)	.28
Self-paced listening (ms)	114 (88)	230 (143)	.20
Cognitive measures			
IQ score (Max = 26)	11.8(2.5)	10.9(3.0)	.03
Backward auditory digit span	5.6(1.2)	5.0(1.4)	.04
Backward visual digit span	5.3(1.3)	5.2(1.5)	n.s.
Forward auditory digit span	6.5(1.1)	5.9(1.2)	.05
Forward visual digit span	6.1(1.2)	5.9(1.4)	n.s.
Non-word recognition task	5.6(1.7)	5.4(1.8)	n.s.

Results

Preliminary Analyses

Table 1 shows the means and standard deviations for native and non-native speakers on all measures as well as effect size estimates (partial η^2) for the differences in mean scores between native and non-native speakers. Not surprisingly, native speakers were always significantly more accurate or significantly faster than non-native speakers on the language measures. Native speakers also scored significantly higher on IQ and the auditory forward and backward digit spans (For further details on performance on the digit span tasks, see Olsthoorn, Andringa, & Hulstijn, 2011), but the sizes of these effects were small. Although a number of skewness and kurtosis values deviated from zero significantly, all absolute values were lower than the maximum values that Kline considers acceptable for SEM analyses, that is, 3 for skewness and 10 for kurtosis (Kline, 2005). In addition, the correlations between all observed variables are presented in Table 2. The table shows rather similar patterns of correlations for the native and non-native speakers, but correlations tend to be slightly stronger for the non-native speakers.

To test whether our predictors of listening comprehension could be considered measures of separable constructs, two multisample first-order CFA were conducted, one for the language measures, and one for the cognitive measures. For the eight language measures, 24 predictors (three subsets for each measure) were submitted to an eight-factor CFA, without imposing any equality constraints between the groups. Overall model fit was good ($\chi^2(448) = 473.7$, $p = .19$, RMSEA = .016, CFI = .996). The highest correlation between the extracted factors was $r = .73$. Thus, there were no reasons to merge factors. For the cognitive measures, a multisample one-factor CFA was conducted. However, this model did not fit. A two-factor model was then tested in which intelligence and working memory capacity were separated. The IQ-test items were randomly divided into three sets to make up three sub scores. This two-factor model did yield good fit ($\chi^2(39) = 30.9$, $p = .81$, RMSEA = .000, CFI = 1.000).

Scale reliabilities were calculated based on the configural model (as presented in the following section). For the native listeners, the reliabilities for vocabulary and segmentation speed were somewhat low (.51 and .66, respectively). For the non-native listeners, this was true for listening, vocabulary, semantic processing speed, grammatical processing speed, and segmentation speed (.66, .42, .56, .55, and .51, respectively). This is not necessarily problematic, as measurement error is partialed out in the extracted factors. All other reliabilities were above .75.

A Structural Model of Listening Comprehension

In the structural model, the aim was to evaluate the overall research question: Which factors explain success in native and non-native listening comprehension? Figure 1 depicts the configural model that was tested. As can be seen, the three accuracy measures (vocabulary, grammatical accuracy, and segmentation accuracy) were regressed onto a Knowledge factor, and the five speed measures (semantic processing speed, grammatical processing speed, segmentation speed, word monitoring, and self-paced listening) were regressed onto a Processing Speed factor. The dependent variable Listening Comprehension—a factor extracted from three sub scores on the discourse comprehension test, which was done to partial out measurement error—was in turn regressed onto the four explanatory factors Knowledge, Processing Speed, Memory, and IQ. As a first step, we tested for configural invariance: Does this configuration hold

Table 2 Pearson correlations for all observed variables (at $\alpha < .05$, 2-tailed) for the native speakers (above the diagonal) and the non-native speakers (below the diagonal)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Listening comprehension	1	.35		.29	−.37	−.36	−.26			.32	.37	.29	.28	.31	
2. Vocabulary	.68	1	.22	.20	−.18		−.21				.27	.19		.27	
3. Grammatical proc. accuracy	.77	.60	1	.23					−.21	.29					
4. Segmentation accuracy	.64	.53	.60	1	−.21				−.19	.23	.25	.21	.27	.31	
5. Semantic processing speed	−.61	−.57	−.52	−.37	1	.63	.35	.24	.30	−.23	−.43	−.41	−.34	−.22	
6. Grammatical proc. speed	−.41	−.41	−.34		.72	1	.42	.29			−.26	−.30			
7. Segmentation speed	−.36	−.37	−.27	−.19	.50	.47	1				−.33	−.33	−.24	−.19	
8. Word monitoring	−.40	−.46	−.42	−.19	.44	.37	.30	1							
9. Self-paced listening	−.40	−.34	−.36	−.27	.50	.30	.41	.34	1				−.22		
10. IQ score	.48	.39	.32	.40	−.39		−.27			1	.26	.20	.22		
11. Backw. auditory digit span	.21	.22		.29	−.20		−.22		−.22	.42	1	.50	.47	.44	.22
12. Backw. visual digit span				.27						.38	.53	1	.47	.46	
13. Forw. auditory digit span	.25			.32	−.21		−.28		−.18	.35	.45	.31	1	.50	
14. Forw. visual digit span	.23			.20			−.22			.34	.43	.57	.40	1	.19
15. Non-word recognition task														.21	1

Note. The numbers in the top row correspond to the numbers in the first column.

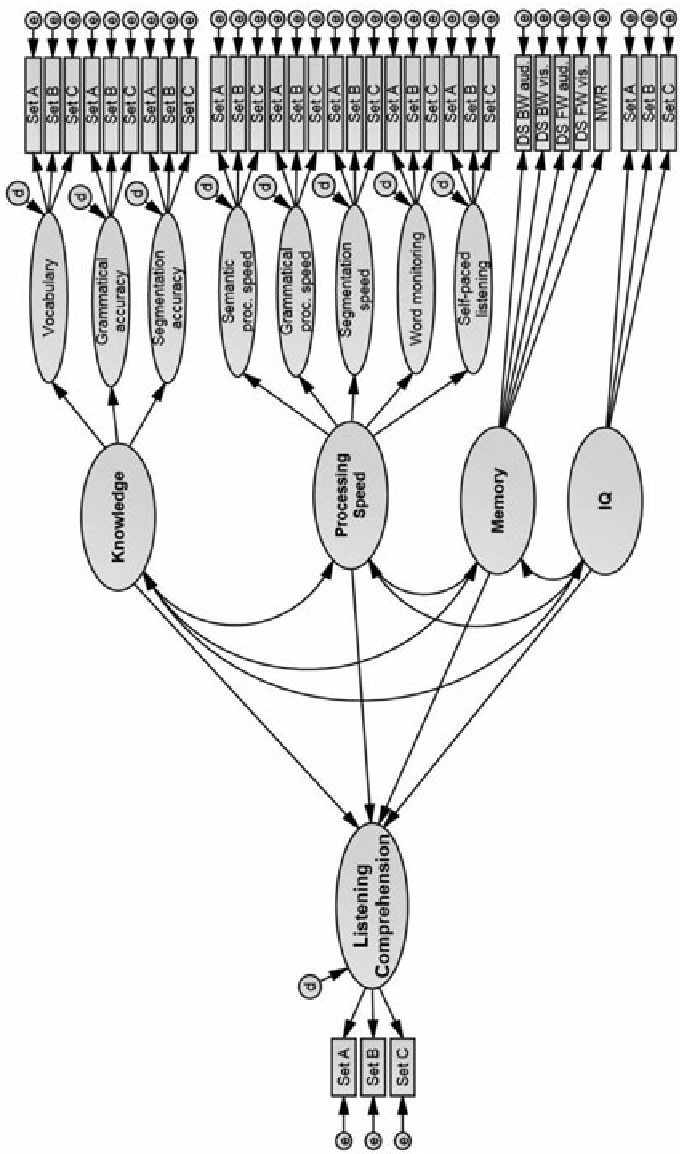

Figure 1 The configural model of listening comprehension. Boxes represent observed variables; ovals represent latent factors. "e" refers to measurement error and "d" to unexplained variance in latent variables.

Table 3 Implied correlations and standardized regression weights of the latent factors with the dependent variable Listening Comprehension for native and non-native listeners as obtained in the final model

	Native listeners		Non-native listeners	
Predictors	Correlations	Stand. regression weights	Correlations	Stand. regression weights
Knowledge	.88	.79	.96	.95
Processing Speed	−.64	−.54	−.67	
Memory	.66		.32	
IQ	.53		.51	.26
Variance explained		91%		96%

for both groups? This required a multisample analysis in which all parameters were allowed to vary between groups. This model fitted the data well ($\chi^2(1084) = 1295.1, p < .001$, RMSEA = .029, CFI = 0.971).

Next, equality constraints were gradually imposed to test for between-group invariance. First, all measurement weights (i.e., the unstandardized regression weights between observed and latent variables) were set equal to test the assumption that the factors extracted from the observed variables were measures of the same constructs for native and non-native listeners. In this model, the error variances were not constrained to be equal. The difference between this model and the configural model was not significant, $\Delta\chi^2(24) = 29.5$, $p = .20$, which confirmed group invariance at measurement level. Then, group invariance at structural level was assessed by setting all structural weights to be equal. This led to a significant loss of model fit, $\Delta\chi^2(10) = 40.5, p < .001$, and the model was rejected. As a last step towards the final and most parsimonious model, structural regression weights that did not differ significantly from zero were set to zero. The resulting model fitted the data well ($\chi^2(1118) = 1344.5$, $p < .001$, RMSEA = .030, CFI = 0.969).

The final model results are displayed in Table 3. For the native listeners, all four predictors correlated substantially with Listening Comprehension. However, the standardized regression weights indicate that neither Memory nor IQ were found to explain any variance beyond what was explained by Knowledge and Processing Speed. Knowledge differences predicted success in Listening Comprehension best, but Processing Speed also explained unique variance. Together, they explained 92 percent of the variance. For the non-native listeners, the picture was somewhat different. Knowledge and Processing Speed also correlated strongly with Listening comprehension; for IQ, a substantial correlation

Table 4 Correlations between the four predictor factors; only significant correlations are given

	Native listeners			Non-native listeners		
	Knowl.	Proc. Sp	Memory	Knowl.	Proc. Sp	Memory
Processing Speed	—			−.68		
Memory	.58	−.45		—	—	
IQ	.48	—	.25	—	—	.58

was also observed, but the correlation for Memory with Listening Comprehension was weak. In the final model, differences in Knowledge explained the bulk of the variance in Listening Comprehension, but there was also a significant contribution of IQ. Together, they explained 96 percent of the variance. Processing Speed and Memory did not explain unique variance in non-native Listening Comprehension. Table 4 presents the correlations between the four predictors for both groups. Interestingly, the pattern of correlations was almost reverse for native and non-native speakers: With the exception of the correlation between IQ and Processing Speed, which was low for both groups, correlations that were relatively strong for the natives were weak for the non-natives and vice versa.

Factor variances and factor latent means were also compared between the groups. The results confirmed the descriptive statistics presented in Table 1. Significantly larger total variances were observed for Listening Comprehension, Knowledge, and Processing Speed for the non-native speakers. As a group, they were more variable in their performance on these factors than the native listeners. Such differences in total variances were not obtained for Working Memory and IQ. The latent means analysis showed significant differences between natives and non-natives for all five factors, the natives always outperforming the non-natives, but the effect was small for working memory and IQ. Thus, the natives performed better than the non-natives and they performed more homogeneously on Listening Comprehension and the Knowledge and Processing Speed factors.

Discussion

The aim of this study was to find out which factors explain success in native and non-native listening comprehension. The results showed that the answer to this question is different for native and non-native speakers. For both, knowledge differences explained variation in discourse comprehension, but for the native speakers Processing Speed also contributed substantially, whereas for the non-native speakers there was a significant contribution of IQ (Table 3). In addition, there were striking differences between the groups in the pattern of correlations observed (Table 4). However, we would like to argue that differences in experience with the Dutch language may explain the observed results for both native and non-native speakers, even though the results for the two groups are different.

For both groups, Knowledge turned out to be the most important predictor of success in listening comprehension. In this study, the Knowledge factor was extracted from measures of vocabulary, grammatical accuracy, and segmentation accuracy. It was made up of the variance these three measures have in common, and it most likely reflected the knowledge required to process speech accurately. For the non-native speakers, it is not surprising that differences in Knowledge explain variation in success in listening comprehension to such a large extent. This finding confirms commonly held beliefs and empirical findings about the role of linguistic knowledge in L2 listening (e.g., Mecartty, 2000; Vandergrift, 2007). However, in Mecartty's (2000) study, measures of knowledge were found to explain a mere 14% of the variance, while they explained large amounts of variance in our sample. This is probably due to the differences in research design between our study and Mecartty's. First, measurement error that attenuates correlations was taken into account in our study due to the factor-analytic design used. Second, the knowledge tests in our study were probably more pertinent to listening; grammatical knowledge was measured in a production task in Mecartty. And third, Mecartty studied university students only. The range of L2 proficiency may have been much larger in our non-native sample; it ranged from B1 to near-native. Consequently, the differences in knowledge of Dutch in our non-native sample will likewise have been rather large. These differences are probably explained by the variables that are usually associated with differences in success of L2 acquisition: They will, to some extent, be explained by the differences and similarities between Dutch and the learner's L1 and talent for language learning, but also by the age at which people started learning, whether people have followed courses of Dutch, whether Dutch is spoken at home and at work, whether they are motivated to engage in Dutch society, and so on. In short, probably much of the variance in L2 Knowledge is explained by the amount of input that people have processed.

Although the variance was significantly smaller, native listeners also exhibited considerable within-group variation in listening comprehension. This is interesting in itself, especially given that the discourse comprehension test we used was designed for non-native speakers. Knowledge was also found to be the best predictor of native listening comprehension. We argue that the explanation for this is the same as for the non-native speakers. For the native speakers too, knowledge differences reflect differences in amount and kind of input processed, in other words, differences in experience. The good listeners may have been the people who are used to processing complex texts and they may have been the more experienced readers and writers. An extra analysis (not reported) was run with Level of Education (LoE) as a dichotomous (high vs. low) predictor, where we assumed people of high education to be more experienced readers and writers. This analysis confirmed our interpretation of the data in that we found a strong correlation between LoE and Knowledge ($r(119) = .65$), which was stronger than the correlations between LoE and the other three factors: People of high LoE had more linguistic knowledge. The relationship between comprehension skill and knowledge of language is most likely reciprocal. More text processing leads to more knowledge, which leads to better comprehension skills.

For the native speakers, success in Listening Comprehension was not just a function of Knowledge; Processing Speed also explained a significant share of the variance. In addition, the correlation between Processing Speed and Knowledge was negative and not significantly different from zero. The negative direction indicates that higher accuracy coincided with faster responses, which suggests there was no accuracy-speed trade-off. This was confirmed by the small or absent correlations between segmentation accuracy and speed, and grammatical accuracy and speed (see Table 2). Taken together this suggests that speed of processing linguistic information is a separate construct that can explain why some native listeners are more successful than others. This may not come as a surprise given findings that comprehension is negatively affected by fast speech (e.g., Wingfield, Peelle, & Grossman, 2003). Our finding suggests that native listeners vary in the ability to deal with the pressure of online speech processing—even if delivered at a normal pace, and that this affects the success of the native comprehension process. It is important to point out that our discourse comprehension task was fairly demanding in this respect. Participants listened to monologues of approximately one minute without the possibility of rewinding, and with limited time to answer the question. For the non-natives, a strong correlation was observed between Processing Speed and Listening Comprehension (Table 3; $r(111) = -.67$), but the standardized regression weights in the same table indicate that Processing Speed did not explain success in Listening beyond what was already explained by Knowledge and IQ. If the correlation between Knowledge and Processing Speed is included in the picture (see Table 4), it is clear that among the non-natives, fast "processors" are also the people who have most linguistic knowledge. It seems that being able to process linguistic information fast and efficiently is a function of the amount of knowledge the non-native listener has; this is not true for the native listeners.

Perhaps one of the most striking results in these data concerns the role of working memory in native speaker discourse comprehension. Previous studies have mostly reported correlations of about $r = .50$ for measures of working memory with discourse comprehension. Despite the fact that digit spans were used rather than listening spans, which are presumably less verbally demanding, a strong correlation (Table 3; $r(119) = .66$) was observed in this study for the native speakers. The strength of this correlation may have been the result of the factor-analytic approach, which led to loss of measurement error in the discourse comprehension and working memory factors (cf. traditional "correction of attenuation"). However, in the presence of Knowledge and Processing Speed, Working Memory did not explain any unique variance in listening comprehension. In addition, Working Memory was found to be correlated substantially with both Knowledge ($r(119) = .58$) and Processing Speed ($r(119) = .45$). This supports the view that digit spans are still measures of verbal working memory, even though they do not require the processing of lengthy sentences (see also Andringa, Olsthoorn, & Hulstijn, 2011; Olsthoorn et al., 2011). For non-native speakers, only marginal correlations were found between Working Memory and Listening Comprehension, Knowledge, and Processing Speed. Working Memory did not explain unique variance in Listening Comprehension.

Taken together, the evidence runs counter to theories that conceive of working memory as a general, language independent cognitive resource (e.g., Just & Carpenter, 1992; King & Just, 1991; MacDonald et al., 1992) and is more in line with theories that equate working memory capacity to experience in language processing (e.g., Ericsson & Kintsch, 1995; MacDonald & Christiansen, 2002). A single-resource account of working memory would predict separable effects of working memory capacity and the language predictors on listening, and it would probably also predict equal or stronger correlations for the non-native speakers on the assumption that listening comprehension would be equally or even more taxing for them. However, stronger correlations for the non-natives were not observed. These results make more sense from the experience-based explanation of working memory. Having much knowledge, being able to process linguistic information quickly, and scoring well on verbal working memory tasks are probably all the result of one's text processing experience, hence the strong correlations for the natives. But why then did we observe a near absence of correlations for Working Memory in non-native listening? This is probably because the non-native memory scores did not reflect experience with the Dutch language. The non-natives may well have used their L1s when doing the memory tasks: If they did, the non-native Working Memory factor may not have been a reflection of their experience with the Dutch language. Our working memory tasks were auditory and visual digit spans, and a non-word recognition task. The visual tasks clearly allow for task execution in the L1. For the auditory tasks (participants listened to the Dutch words for digits), one cannot be certain: participants may have translated to their L1s. Similarly, the non-word recognition task consisted of pseudowords, which makes it unlikely that the score on this task will have been a reflection of experience with Dutch. Indeed, the non-word recognition task made the smallest contribution to the working memory factor. In short, the working memory factor that was extracted from the memory scores probably represented language experience for the natives, but not for the non-natives.

A similar point can be raised about the role of IQ in Listening Comprehension. Despite a substantial correlation of $r(119) = .53$ between IQ and Listening Comprehension, IQ was not found to predict unique variance in native speaker listening. However, IQ was substantially correlated with Knowledge ($r(119) = .48$). This can be taken to mean two things. It could mean that reasoning ability facilitates the ability to accurately interpret sentence meanings. This seems unlikely, as reasoning ability is associated with consciously controlled reasoning processes, whereas lower-level sentence processes (as measured by the tasks from which our Knowledge factor was extracted) are thought to be implicit and automatic, at least for native speakers. Alternatively, the correlation between Knowledge and IQ could also be the result of the fact that people of higher IQ tend to have higher levels of education and better jobs. As a result, they have more experience in language processing and more knowledge that is relevant to explaining success in listening. For the non-native speakers, the observed correlation between IQ and listening was also substantial $r(111) = .51$, but it did not correlate with Knowledge and Processing Speed. IQ did survive in the regression analysis as a predictor of variance in discourse comprehension. This suggests that non-native listeners make use of meta-cognitive reasoning abilities as measured by the complex matrices component of the WAIS-III (Wechsler, 1997) during discourse comprehension. This is perhaps not surprising: listening is likely more demanding for non-native speakers.

The goal of this study was to fill some of the voids in listening research. In the introduction, we pointed to the fact that the vast majority of research conducted on native listening comprehension has been experimental in nature.

In the domain of non-native language processing, differential approaches have been more common, but have seldom been applied to investigate the construct of listening. As early as 1957, Cronbach lamented the division between differential and experimental studies of behavior: "The correlational psychologist is in love with just those variables the experimenter left home to forget" (p. 674). He expressed hopes for the development of techniques that would allow for the combination of both approaches (Cronbach, 1957). Perhaps this study, which was still observational in nature, is a step in this direction in that it exemplifies a statistical technique that may be extended to more experimental designs. This study may also exemplify the added value of a multivariate perspective in research designs. Its multivariate design is probably this study's biggest strength because it allowed for the assessment of the importance of variables in conjunction. This way, we observed that working memory capacity defined as an independent general cognitive function—a concept that features prominently in many theories of comprehension, does not explain why some are better listeners than others. Native listening proficiency is a function of both linguistic knowledge and efficiency of processing linguistic information, whereas non-native listening proficiency is mostly linguistic knowledge and a little bit of reasoning ability.

References

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

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Andringa, S. J., Olsthoorn, N. M., & Hulstijn, J. H. (2011). *Sources of individual differences in native speaker auditory sentence processing*. Unpublished manuscript.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database*. Philadelphia: Linguistic Data Consortium, University of Philadelphia.

Boersma, P., & Weenink, D. (2005). *PRAAT*. Version 5.0.35. Retrieved October 6, 2008, from http://www.praat.org

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York: Routledge.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*, 77–126.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.

Coussé, E., Gillis, S., & Kloots, H. (2007). Verkort, verdoft, verdwenen: Vocaalreductie in het corpus gesproken nederlands. *Nederlandse Taalkunde*, *12*(2), 109–138.
Cronbach, L. J. (1957). The two disciplines of scientific psychology. *The American Psychologist*, *12*(11), 671–684.
Cutler, A., & Clifton, C. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford, UK: Oxford University Press.
Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 113–121.
Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behaviour*, *19*, 450–466.
Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 561–584.
Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*, 1–18.
Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*, 422–433.
De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking. *Studies in Second Language Acquisition*, *34*, 5–34.
Dirksen, A. (2008). *Fluency TTS*. Amsterdam: Fluency.
Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.
Ernestus, M. T. C. (2000). *Voice assimilation and segment reduction in casual dutch: A corpus-based study of the phonology-phonetics interface*. Utrecht, Netherlands: LOT.
Fallon, M., Peelle, J. E., & Wingfield, A. (2006). Spoken sentence processing in young and older adults modulated by task demands: Evidence from self-paced listening. *Journal of Gerontology*, *61B*, 10–17.
Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*, 491–505.
Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A., & McFarlane, D. K. (1996). Effects of lexical frequency and syntactic complexity in spoken-language processing: Evidence from the auditory moving-window technique. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 324–335.
Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System*, *32*, 363–377.
Fleishman, J., & Benson, J. (1987). Using LISREL to evaluate measurement models and scale reliability. *Educational and Psychological Measurement*, *47*, 925–939.
Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*, 25–38.

Havik, E., Roberts, L., Van Hout, R., Schreuder, R., & Haverkort, M. (2009). Processing subject-object ambiguities in the L2: A self-paced reading study with german L2 learners of dutch. *Language Learning*, *59*, 73–112.

Hazenberg, S., & Hulstijn, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics*, *17*, 145–163.

Het corpus gesproken nederlands (2006). Leiden: Instituut voor Nederlandse Lexicografie.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Coventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Jincho, N., Namiki, H., & Mazuka, R. (2008). Effects of verbal working memory and cumulative linguistic knowledge on reading comprehension. *Japanese Psychological Research*, *50*, 12–23.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension. *Psychological Review*, *99*, 122–149.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*, 580–602.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford.

Kloots, H., De Schutter, G., Gillis, S., & Swerts, M. (2003). Vocaalreductie in spontaan gesproken Standaardnederlands: Een verkennende studie. In T. Koole, J. Nortier, & B. Tahitu (Eds.), *Artikelen van de vierde sociolinguïstische conferentie* (pp. 224–233). Delft, Netherlands: Eburon.

Koster, C. J. (1987). *Word recognition in foreign and native language: Effects of context and assimilation*. Dordrecht, Netherlands: Foris.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review*, *109*, 35–54.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, *24*, 56–98.

Mack, M. (1988). Sentence processing by non-native speakers of English: Evidence from the perception of natural and computer-generated anomalous L2 sentences. *Journal of Neurolinguistics*, *3*, 293–316.

MacWhinney, B. (2008). A unified model. In P. Robinson, & N. Ellis (Eds.), *The handbook of cognitive linguistics and second language acquisition* (pp. 341–371). New York: Routledge.

Masson, M. E., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology*, *75*(2), 314–318.

McClelland, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

Mecartty, F. H. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning*, *11*, 323–348.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, *34*, 191–243.

Olsthoorn, N. M., Andringa, S. J., & Hulstijn, J. H. (2011). *Visual and auditory digit span performance in native and non-native speakers*. Unpublished manuscript.

Papadopoulou, D., & Clahsen, H. (2003). Parsing strategies in LI and L2 sentence processing: A study of relative clause attachment in Greek. *Studies in Second Language Acquisition*, *25*, 501–528.

Payne, T. W., Kalibatseva, Z., & Jungers, M. K. (2009). Does domain experience compensate for working memory capacity in second language reading comprehension? *Learning and Individual Differences*, *19*, 119–123.

Pearlmutter, N. J., & MacDonald, M. C. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, *34*, 521–542.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*, 1–23.

Roberts, L. (2012). Individual differences in second language sentence processing. *Language Learning*, *62*, 172–187.

Sanders, L. D., Neville, H. J., & Woldorff, M. G. (2002). Speech segmentation by native and non-native speakers: The use of lexical, syntactic, and stress-pattern cues. *Journal of Speech*, *45*, 519–530.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime reference guide*. Pittsburgh, PA: Psychology Software Tools.

Schoonen, R., Hulstijn, J., & Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8 and 10. *Language Learning*, *48*, 71–106.

Schoonen, R., Van Gelderen, A., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P., et at. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, *53*, 165–202.

Schoonen, R., Van Gelderen, A., Stoel, R. D., Hulstijn, J., & De Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning*, *61*, 31–79.

Townsend, D. J., & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*, 204–224.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

Van Gelderen, A., Schoonen, R., De Glopper, K., Hulstijn, J. H., Simis, A., Snellings, P., et al. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first-and second-language reading comprehension: A componential analysis. *The Journal of Educational Psychology*, *96*, 19–30.

Van Hell, J., G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research*, *26*, 43–74.

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, *40*, 191–210.

Vandergrift, L., Goh, C. C. M., Mareschal, C. J., & Tafaghodtari, M. H. (2006). The metacognitive awareness listening questionnaire: Development and validation. *Language Learning*, *56*, 431–462.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter. *Psychological Review*, *103*, 761–772.

Waters, G. S., & Caplan, D. (2004). Verbal working memory and on-line syntactic processing: Evidence from self-paced listening. *The Quarterly Journal of Experimental Psychology*, *57A*, 129–163.

Waters, G. S., Caplan, D., & Yampolski, S. (2003). On-line syntactic processing under concurrent memory load. *Psychonomic Bulletin & Review*, *10*, 88–95.

Wechsler, D. (1997). *The wechsler memory Scale–III manual*. San Antonio, TX: The Psychological Corporation.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250–271.

Wingfield, A., Peelle, J. E., & Grossman, M. (2003). Speech rate and syntactic complexity as multiplicative factors in speech comprehension by young and older adults. *Neuropsychology, Development, and Cognition*, *10*, 310–322.

Wolff, D. (1987). Some assumptions about second language text comprehension. *Studies in Second Language Acquisition*, *9*, 307–326.