The effects of age and level of education on the ability of adult native speakers of Dutch to segment speech into words

Hulstijn, J.H.; Andringa, S.J.

Published in:
Above and beyond the segments: experimental linguistics and phonetics

DOI:
10.1075/z.189

The effects of age and level of education on the ability of adult native speakers of Dutch to segment speech into words

Jan H. Hulstijn and Sible Andringa

This study tested Hulstijn's (2011) hypothesis that adult native speakers share the ability to process every-day speech, although older people do so more slowly than younger people. In two segmentation tasks, segments of speech were presented consisting of two to four highly common words. In the Count Task, participants decided as quickly as possible how many words they had heard. In the Dictation Task, participants wrote down what they had heard. Participants (N = 235) were selected on the basis of their age (young vs senior) and level of education (low vs high). Some of the findings are at variance with the predictions, which may or may not have been caused by a task confound.

Introduction

As in many other sciences, in the discipline of linguistics researchers can be found mainly working on the question of what people have in common and, in contrast, researchers who devote most of their work on individual differences. Thus, while much work of syntacticians aims at investigating whether or not all adult native speakers of a language share a common underlying language competence (e.g., a universal grammar of some sort) and while most work in psycholinguistics aims at uncovering universal principles in the processing of linguistic information, researchers working in other sub-disciplines, such as sociolinguistics and clinical linguistics, focus primarily on individual differences. In phonetics the study of what is common to all speakers of a language goes hand in hand with the study of how and to what extent the speakers of a language differ. Similarly, in the study of first language acquisition, it is important to establish which elements of the language are acquired by all (typically developing) children and adolescents, and which elements are acquired by some but not all native speakers. Once we know what is acquired by all native speakers and what is acquired by some, we know which phenomena a theory of language acquisition has to explain.

It is obvious that people differ in an incredibly large number of physical and mental respects and this is also true with respect to the knowledge they acquire of their first language. One might perhaps initially be inclined to conceptualize individual differences in the knowledge or control of language skills as an extended oval along a fairly straight regression line in which all native speakers can be neatly placed, with most speakers placed in the area where the oval is widest, as in a normal distribution, with two equally long tails at either side of the mean. Alternatively, one might think of individual differences in terms of a Zipfian distribution, with some elements of language acquired by all native speakers and other elements acquired by some native speakers, depending on language-use related factors such as literacy, level of education, profession, leisure-time activities and age. Using ideas proposed in sociolinguistics (Bernstein, 1971), child cognitive development (Donaldson, 1978; Bialystok, 1986), and educational bilingualism (Cummins, 1980a, 1980b), Hulstijn (2007, 2011, in progress) makes a distinction between Basic Language Cognition (BLC) and Higher Language Cognition (HLC) (also called Extended Language Cognition). Whereas BLC is shared by all adult native speakers (not affected by language-related disorders), HLC is the part of knowledge and control of language where individual differences can be observed. In Hulstijn (2011, in progress) a number of hypotheses are derived from the BLC-HLC distinction. Of these hypotheses, the following two are relevant in the context of the study presented in this paper.

H1. All adult L1ers (not suffering from mental disorders), regardless of differences in age and intellectual functioning, are able to comprehend and produce, both correctly and quickly, isolated utterances consisting of high-frequency lexical phrases and high-frequency morphosyntactic structures, when these utterances are perceived under normal acoustical conditions. (Hulstijn, 2011, p. 231)

H2. Although the speed with which humans can process information increases over time until it reaches a peak around the age of 22 and from the age of 27 on gradually decreases (Salthouse, 2009), the vast majority of old people remain capable of processing linguistic information fast enough to allow for relatively unimpaired functional language use, provided that they continue to practice their language skills on a daily basis and do not suffer from severe mental disorders. (Hulstijn, 2011, p. 232)

As Hulstijn (in progress, Chapter 6) argues, theories and hypotheses should ideally be both somewhat plausible and somewhat implausible so that researchers will be motivated to invest time and effort to falsify them, thereby increasing our understanding of the phenomena we want to explain. In this spirit, we examine in the study reported here whether or not evidence for H1 and H2 could be found in the skill of segmenting speech into words (see Cutler, 2012, for a review of the segmentation literature). The participants in this study heard short fragments of speech (in the Dutch language) such as *dat weet* 'that know' or *in ieder geval fijn* 'in any case nice' and had to tell how many words they had heard (by pressing the appropriate number key on a computer keyboard) in the first task or write down which words they believed they had heard in the second task. Participants differed in their age and level of education. The speech fragments only contained common words, with which all native speakers were assumed to be familiar. Thus, being able to correctly segment such fragments was assumed to belong to BLC and participants' performance should not be moderated by their age or level of education (H1). However, older people may need more time to process the stimuli and we therefore expected to find an age effect in the speed with which participants made their decisions about the number of words they had heard, with older participants needing more time than younger participants (H2).

Method

This paper reports on data collected in the framework of a larger, NWO-funded project, called *Studies in Listening* (StiLis), in which 345 people took part (235 native speakers [NS] and 110 non-native speakers of Dutch), who were tested on a large number of tests (Andringa, Olsthoorn, Van Beuningen, Schoonen, & Hulstijn, 2012).[1] The present study is restricted to an examination of the performance of the participants in the speech-segmentation task as affected, on the one hand, by participant variables (age, level of education [LoE], nonverbal intelligence, working-memory capacity [digit span], and reaction time in a nonverbal task) and on the other hand by stimulus characteristics (stimulus length and reduction of articulation).

Participants

The participants were recruited through advertisements posted in several educational institutes, supermarkets, community centres and through networks of relatives and friends. We recruited 235 NSs on the basis of their age (young vs senior) and LoE (low vs high) so that four groups of approximately 60 subjects could be formed by a factorial crossing of Age and LoE. The young ($n = 119$; 82 female, 37 male) and senior subjects ($n = 116$; 78 female, 38 male) were between 19–40 and 56–82 years old, respectively. Subjects with low LoE ($n = 117$; 81 female, 36 male) were actively enrolled in or had finished a vocational education; subjects with high LoE ($n = 118$; 79 female, 39 male) were actively enrolled in or had completed an education at college level or higher. All participants were financially compensated. Subjects with hearing problems and people using medication that might impair their ability to perform reaction-time

1. The official title of the project is *Towards a theory of second-language proficiency* (NWO grant 36-70-230), awarded to Jan Hulstijn.

tasks were excluded from the study.[2] Subjects performed a range of tasks, many of which not relevant for the present study, in two sessions of two hours on separate days, with short breaks between tasks. All subjects signed a consent form.

Tasks and materials

Segmentation

The segmentation task was designed to test to what extent subjects were able to recognize common words in normal speech. The task comprised two parts, each introduced by instructions and a practice session. We will refer to these two parts of the segmentation task with the labels *Count Task* and *Dictation Task*, respectively. Hearing ability (hearing loss) was not independently assessed. However, all participants had reported that they did not suffer from hearing problems and they performed the segmentation tasks while wearing headphones and after they had been given ample opportunity to adjust the volume of the speakers.

The Count Task

The Count Task consisted of five practice trials and 45 experimental trials in random order. In part one, each trial played back a recorded utterance fragment and prompted subjects to enter the number of words they thought they had heard, by choosing a number from 1 to 5 on a keyboard. Subjects were instructed to respond as quickly and accurately as possible, the speed and accuracy of their responses being recorded for analysis. The interval between a response and the presentation of the next stimulus was 2 seconds.

The Dictation Task

In the Dictation Task, subjects were first given three practice trials and then heard the same 45 stimuli of the Count Task, but presented in a different random order. They typed the words they thought they had heard, as in a dictation. They could do so at their own pace and reaction times were not recorded. In scoring the responses, minor spelling errors were ignored as long as it was clear without doubt whether the stimulus words had been correctly identified; one point was given for each correct response.

---

2. The focus of the project (Andringa et al., 2012) of which the study reported in this paper is a minor part, was concerned with the associations between higher-order listening comprehension on the one hand and, on the other hand, lexical knowledge, accuracy and speed of verbal-information processing, working-memory capacity and non-verbal intelligence. This is the main reason why no hearing test was administered, given that subjects were using headsets and were allowed to adjust the volume to their convenience in all listening tasks. In hindsight, we acknowledge it would have been better, had a hearing test been administered.

Thus, while the Count Task allowed us to examine both accuracy and speed, the Dictation Task allowed us to examine accuracy only. However, while, in the Count Task, subjects could guess the number of words without being able to tell exactly which words they had heard, this was impossible in the Dictation Task. Therefore, the accuracy scores in the Dictation Task might give us a more valid measure of segmentation skills than those in part one.

Stimulus materials

The 45 stimuli consisted of strings of two, three, or four words (although, in the Count Task, participants were told to choose from 1 to 5), ranging from two syllables (i.e., two one-syllable words) to six syllables. The stimuli represented legal word strings that could occur anywhere in an utterance. All words were common lexical items that all participants could be assumed to be familiar with. Some examples are: *bij mij* 'with me', *over de* 'about the', *moet je* 'should you', *is ongeveer wat* 'is approximately what', *gegeven moment was* 'given moment was'. All stimuli with their characteristics are listed in Appendix 1 and 2.

All 45 stimuli were spoken in standard Dutch by the same female native speaker and articulated as in normal connected speech addressed to a hearer at around one meter distance, without any word or words stressed, in a relatively flat intonation. Thus, no pauses existed between words; the words were articulated in a concatenated manner, with normal between-word assimilations. The stimuli were recorded in a sound proof studio at 16 bit, 44.1 kHz. Half of the stimuli (23 items) were pronounced with reductions characteristic of standard colloquial speech (Appendix 2). For instance, the stimulus *ik heb het* (three one-syllable words) was spoken as *kepət* (two syllables with the vowel of the second syllable reduced to a schwa). The other stimuli (22 items) were spoken without reductions (Appendix 1).

The stimuli were selected from an original set of 60 items, used in a pilot study, conducted by Meijer (2008). Meijer asked 15 native and 14 non-native speakers of Dutch to perform both tasks. Two stimulus lists (A and B) were made, such that half of the stimuli occurred in their normal full form in A and in their reduced form in B and that the other half of the stimuli occurred in the reduced form in A and in their normal full form in B. The main purpose of the pilot study was to explore the feasibility of the two tasks and to establish whether any stimuli were too difficult for the native speakers or too easy for the non-native speakers. Because, in the main study a large battery of tests had to be administered and the administration of the segmentation task was not supposed to consume much time, only 45 stimuli were presented in the Count and Dictation tasks, 22 in the normal full form and 23 in the reduced form. The pilot study had shown that in this limited form, the Count and Dictation tasks could produce valid and reliable information about individual differences in the ability of native and non-native speakers to segment short fragments of speech, with minimal or no syntactic or pragmatic information. Thus, in the data reported here, articulatory reduction of the stimuli (absent or present) was *not* a factor in the study's design.

Control tasks
An audio-visual *nonverbal reaction-speed* task was included to be able to control for differences between participants in the speed with which they could press the response keys in a number of the project's language-processing tasks, including the segmentation tasks. To control for *working-memory capacity* for aural verbal information, participants performed an aural backward digit-span task. They heard series consisting of 2 to 8 digits, spoken in Dutch, and reproduced them by typing in the reverse order. To control for individual differences in *reasoning ability*, intelligence was assessed with the complex matrices component of the WAIS-III, a measure of nonverbal IQ (Wechsler, 1997). For details about these three control measures, see Andringa et al. (2012).

Data handling and statistical analyses

The first step in the analyses was to inspect each measure for outliers. Reaction time (RT) 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 age groups separately. This never constituted more than 2.4% of the data points in any task. In addition, for the speed-of-language-processing tasks (including the Count Task), latencies to inaccurate responses were considered invalid and were also set to missing (12.5% in the Count Task). All missing values were then imputed by means of the full information maximum likelihood estimation procedure in SPSS, and mean RTs per subject were calculated.

To determine the effects of Age and LoE on the accuracy and speed with which subjects performed the segmentation tasks, linear mixed effects modelling (LMM) was used. The advantage of using linear mixed modelling is that there is no need to aggregate responses to individual items across item types or participants. Instead, multilevel techniques model these dependencies, thus adjusting for correlations between cases due to these so-called random or contextual variables. Participants and items were jointly modelled as random effects, allowing us to assess to what extent the variance present in the data was explained by the variables operationalized at item level (articulation reduction and stimulus duration) and at participant level (Age, LoE, and the three control variables: nonverbal reaction speed, working-memory capacity, and nonverbal IQ) (Baayen, Davidson, & Bates, 2008). We were also able to look at cross-level interactions between sentence and participant level variables.

Results

In this section we first present the results of the LMM analyses and then proceed with an inspection of the difficulty of individual stimuli.

Dictation Task (response accuracy)

Response-accuracy data in the Dictation Task were obtained for 235 subjects (henceforth Ss). As can be seen in Table 1, senior and LoE-low Ss performed, on average, less well than younger and LoE-high Ss, respectively. An LMM analysis was conducted on the accuracy data in the Dictation Task with the following independent factors: Age (young vs senior), LoE (low vs high), Reduction (normal vs reduced articulation), and Stimulus Duration. The interactions between Age, LoE and Reduction were also included. Significant main effects were obtained for Age ($\beta = -2.16$ [$SE = 0.41$], $p < 0.001$), LoE ($\beta = 0.47$ [$SE = 0.15$], $p < 0.001$), Reduction ($\beta = -2.29$ [$SE = 0.49$], $p < 0.001$), and Stimulus Duration ($\beta = -0.003$ [$SE = 0.001$], $p < 0.001$). Correct responses were less likely for the senior and the lower educated Ss. This was also true when items were reduced or relatively short in duration. Interestingly, an interaction was found between Age and Reduction, which indicated that the negative effects of reduction were smaller for the senior Ss ($\beta = .62$ [$SE = .21$], $p = 0.003$). Next, additional analyses were run for each of the control variables in which main effects as well as their interactions with Age, LoE and Reduction were included. When Working Memory (WM) was introduced, the model did not change, but an additional main effect was observed for WM ($\beta = 0.15$ [$SE = 0.06$], $p = 0.01$), indicating that the likelihood of a correct response in the Dictation Task was larger when Ss had more working-memory capacity. No effects were found for IQ or non-verbal reaction speed.

Count Task (response accuracy)

When the Count Task was used as dependent variable, the results changed. There was no longer an effect for Age, and the interaction between Age and Reduction also disappeared. Significant effects were found for LoE ($\beta = 0.27$ [$SE = 0.09$], $p = 0.005$), Reduction ($\beta = -1.37$ [$SE = 0.24$], $p < 0.001$), and Duration ($\beta = 0.003$ [$SE = 0.001$], $p < 0.001$. Correct responses were less likely for the lower educated participants, and for responses to reduced and relatively short items. The additional analyses for the control variables showed main effects of IQ ($\beta = 0.05$ [$SE = 0.02$], $p = 0.014$) and WM ($\beta = 0.08$ [$SE = 0.03$], $p = 0.025$), such that Ss with higher scores on these tasks were more likely to provide correct answers.

Table 1. Response Accuracy (Max = 45) in the Dictation and Count Tasks and Response Speed (ms) in the Count Task, by Age and Level of Education (LoE).

	N	**Dictation Accuracy**		**Count Accuracy**		**Count RT (ms)**	
		M	*SD*	*M*	*SD*	*M*	*SD*
Young							
LoE-low	58	40.4	1.9	38.6	4.4	1452	442
LoE-high	61	41.5	1.9	40.0	3.0	1219	340
Total Young	119	41.0	2.0	39.3	3.8	1333	409
Senior							
LoE-low	59	36.6	6.5	38.9	3.5	1842	495
LoE-high	57	38.4	5.5	40.1	2.3	1629	409
Total Senior	116	37.4	6.1	39.5	3.0	1737	465
LoE-low	117	38.5	5.2	38.8	4.0	1649	507
LoE-high	118	40.0	4.4	40.0	2.7	1417	426
All subjects	235	39.2	4.8	39.4	3.4	1535	482

Count Task (speed)

When response times were used as dependent variable, effects were found for Age ($\beta = 413.5$ [$SE = 58.2$], $p < 0.001$), LoE ($\beta = -278.2.5$ [$SE = 65.2$], $p < 0.001$), Reduction ($\beta = 436.4$ [$SE = 140.9$], $p < 0.001$) and Duration ($\beta = 1.45$ [$SE = 0.32$], $p < 0.001$). In addition, an interaction was observed between LoE and Reduction ($\beta = 40.3$ [$SE = 20.1$], $p = 0.045$). Ss were faster when they were young or highly educated, and slower when items were reduced or relatively long in duration. The interaction indicates that while the highly educated Ss are faster, they also slowed down more when item articulation was reduced. The analyses for the control variables yielded main effects for WM ($\beta = 94.8$ [$SE = 20.4$], $p < 0.001$) and non-verbal reaction speed ($\beta = 69.5$ [$SE = 31.2$], $p = 0.024$), but there were no interactions with Reduction, Age and LoE. Ss tended to be faster on the speed task when they scored higher on WM or when they were faster in the nonverbal reaction task.

Analyses of errors in the Dictation Task

In this section we present analyses of the difficulty of the task of segmenting the stimuli into the correct words. Table 2 shows the error percentages in the Dictation Task.

No significant associations were obtained between error incidence rates and stimulus length in terms of number of words or syllables or in terms of reduction gravity.

For all items, it was the case that some responses clearly resulted from unintended typing errors (e.g., *gzeg maar* instead of *zeg maar*). Some resulted from spelling

Table 2. Percentages of errors in the Dictation Task, by Stimulus type and Participant group.

Stimulus type	Number of stimuli	Young subjects (N = 119)	Senior subjects (N = 116)	All subjects (N = 235)
Full articulation	22	1.6%	7.7%	4.7%
Reduced articulation	23	15.9%	25.6%	20.7%

ignorance or lack of spelling monitoring probably because of time pressure (e.g., *behorlijk veel* instead of *behoorlijk veel*), or substandard varieties of the language (e.g., *bij mijn* instead of *bij mij*). Cases of separate words spelled as one word could result in ambiguous responses (e.g., *moetje* instead of *moet je*). In many cases it was impossible to decide whether the stimulus words were correctly identified, which may have been partly due to the intricacies of the Dutch spelling rules (e.g., *wil je darna met* instead of *wil je daarna met*). Some 'errors' consisted of adding words not present in the stimulus (e.g., *het is niet duidelijk wie* instead of *is niet duidelijk wie*; *hij heeft namelijk veel* instead of *heeft namelijk veel*). We coded these responses as correct because all stimulus words were identified correctly. A large portion of the errors could be interpreted as genuine segmentation errors, demonstrating the validity of the test items (assuming that no typing infelicities were involved). Examples of such genuine segmentation and identification errors were *vorig genoeg* instead of *vorige nog*; *en ieder geval fijn* instead of *in ieder geval fijn*. An example of the difficulty to decide whether a response resulted from misperception or from a spelling/typing infelicity is *het ziet precies op* instead of *het zit precies op*; *ik had gewoon niets* instead of *ik had gewoon iets*.

Some of the items whose stimulus was articulated with the reduction or omission of one or several phonemes appeared to lead to unwanted ambiguity. For example, the omission of /t/ in *die lijk me* for *die lijkt me*, produced 19 *die lijk me* responses, which could not be judged to be truly incorrect (because *lijk* is a Dutch word). Similarly, the reduction of *is ongeveer wat* to *is oveer wat* produced 29 *is of er wat* responses and 9 *is of weer wat* responses. While these two responses might not be taken as incorrect renderings of the acoustic stimulus, the responses *is ook weer wat*, *is over wat*, and *is zo veel wat* were incorrect renderings. Thus, this item was partially infelicitous and partially felicitous.

Some other items with reductions did not lead to ambiguity. For example, some Ss failed to identify *to wel* (/tɔʋɛl/) as a reduced instance of *toch wel*. Similarly, the reduction of *ga ik ook* to *gaak ook* produced five incorrect *gaak ook* responses. Interestingly, omission of the final /t/ in the first word of *moet je anders doen* to *moe je anders doen* produced hardly any errors with respect to the identification of *moet* as the first word but many errors in perceiving *anders* as *alles*. The omission of the final /t/ of the word *vast* in the stimulus *en vast zit*, reduced to *en vas zit* (/ɛfasɪt/) produced errors not only with respect to the recognition of *vast zit* (e.g., erroneously perceived as *was het*), as expected, but also with respect to the first word (erroneously perceived as *aan*, *en* and *het*).

Extreme cases of reduction, such as *ik weet eik ook* for *ik weet eigenlijk ook* and *of het mook is* for *of het mogelijk is*, lead to large numbers of failed recognition. The item in which *gegeven moment was* reduced to *geen moment was* was obviously invalid because *geen moment* is correct Dutch as well.

Correlations within and between tasks

Dictation-Task items were easy or difficult to young and senior Ss to approximately the same extent, with $r = .81$ ($N = 45$ stimuli), computed from the errors produced by all Ss. Accuracy scores of senior Ss in the Count Task correlated substantially with those in the Dictation task, with $r = .91$. The correlation was much lower in the case of the young Ss ($r = .71$) because many young Ss performed at ceiling.

Discussion

Summary of the findings

On the basis of H1, we expected not to find main effects of Age and LoE because the segmentation task consisted of stimuli that were assumed to belong to normal every-day aural-language input to all adult native speakers. Although performance accuracy was generally very high (Table 1), senior and LoE-low participants performed slightly but statistically significantly less well than young and LoE-high participants. Thus it is debatable whether H1 was supported or falsified. H2 predicted that senior participants would perform more slowly than young participants. Support for this prediction was found (Table 1) and this finding reliably replicates what has been found by others (see Mulder & Hulstijn, 2011, for a literature review). However, an unpredicted main effect of LoE was also obtained.

Interpretation of the findings

The Count and Dictation tasks are behavioral tasks in the sense that subjects not only covertly process the auditory stimuli, as in the processing of speech in every-day communication, but subjects also perform an act of overt behavior, namely pressing a number key (Count Task) or writing down the words perceived (Dictation Task). These acts require cognitive processes of an executive type (counting, spelling), in addition to the word recognition (segmentation and identification) processes themselves. Thus, one might argue that the behavioral measures in this study reflect the sum of several types of processing rather than segmentation and identification alone, even after controlling for the effects of (i) working-memory capacity, (ii) reasoning ability and (iii) reaction-speed in a nonverbal task. This confound may explain why effects of Age and LoE were found in the Dictation task and why LoE effects were found in the Count Task, at variance with H1 and H2. It seems, therefore, that the findings of this study have not falsified H1 (and H2 to a lesser extent) unequivocally and that the jury is still out. Follow-up studies measuring participants' ERPs in processing simple speech, in tasks not requiring selective decisions of a metacognitive kind, might resolve this issue.

Acknowledgement

The first author of this paper would like to acknowledge that he had the great pleasure of working with Vincent van Heuven in a NWO-funded research project more than fifteen years ago, thereby learning a great deal from Vincent van Heuven.

References

Andringa, S., Olsthoorn, N., Van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning*, 62[Supplement], 49–78. DOI: 10.1111/j.1467-9922.2012.00706.x

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. DOI: 10.1016/j.jml.2007.12.005

Bernstein, B. (1971). *Class, codes and control. Vol. 1: Theoretical studies towards a sociology of language*. London, UK: Routledge & Kegan Paul.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development*, 57, 498–510. DOI: 10.2307/1130604

Cummins, J. (1980a). The construct of language proficiency in bilingual education. In J. E. Alatis (Ed.), *Current issues in bilingual education. Georgetown University Round Table on Languages and Linguistics (GURT) 1980* (pp. 81–103). Washington: Georgetown University Press.

Cummins, J. (1980b). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly*, 14, 175–187. DOI: 10.2307/3586312

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: The MIT Press.

Donaldson, M. (1978). *Children's minds*. Glasgow, UK: Collins.

Hulstijn, J. H. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91, 663–667. DOI: 10.1111/j.1540-4781.2007.00627_5.x

Hulstijn, J. H. (2011). Language proficiency in native and non-native speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, 8, 229–249. DOI: 10.1080/15434303.2011.565844
Hulstijn, J. H. (in progress). *Language proficiency in native and non-native speakers: Theory, research and measurement*.
Meijer, N. (2008). *Segmentation of reduced speech by native and non-native listeners of Dutch*. MA thesis in Linguistics, University of Amsterdam.
Mulder, K., & Hulstijn, J. H. (2011). Linguistic skills of adult native speakers, as a function of age and level of education. *Applied Linguistics*, 32, 475–494. DOI: 10.1093/applin/amr016
Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of aging*, 30, 507–514. DOI: 10.1016/j.neurobiolaging.2008.09.023
Wechsler, D. (1997). *WAIS-III administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Appendix 1

The 22 stimuli articulated without reduction or omission of phonemes

Item	Stimulus	N of words	N of syllables
11	wat er gezegd is	4	5
12	heeft namelijk veel	3	5
13	mijn plan	2	2
14	bedoel ik niet	3	4
15	laat ze altijd gaan	4	5
16	erg makkelijk voor	3	5
17	vorige nog	2	4
33	dat weet	2	2
34	het zit precies op	4	5
35	zeg maar	2	2
36	dan gaan ze	3	3
37	wil je daarna met	4	5
38	dat vind ik	3	3
40	moet je	2	2
51	is natuurlijk al	3	3
52	is niet duidelijk wie	4	6
54	komt redelijk vaak	3	5
55	behoorlijk veel	2	4
56	te persoonlijk om	3	5
57	bepaalde tijd	2	4
58	ik had gewoon iets	4	5
60	in ieder geval fijn	4	5

Appendix 2

The 23 stimuli articulated with reduction or omission of phonemes

Item	Stimulus	Reduction type	Correct response	N of words	N of syllables*
1	bə mij	Vowel reduction	bij mij	2	2 (3)
3	t hek	Vowel omission	het hek	2	1 (2)
5	hij s lang	Vowel omission	hij is lang	3	2 (3)
8	ga k ook	Vowel omission	ga ik ook	3	2 (3)
9	k was ontzettend blij	Vowel omission	ik was ontzettend blij	4	5 (6)
10	vloren heeft	Vowel omission	verloren heeft	2	3 (4)
44	hoeft dan almaal niet	Vowel omission	hoeft dan allemaal niet	4	5 (6)
21	en vas zit	Consonant omission	en vast zit	3	3 (3)
22	je moet eers gaan	Consonant omission	je moet eerst gaan	4	4 (4)
23	denk men	Consonant omission	denkt men	2	2 (2)
24	die lijk me	Consonant omission	die lijkt me	3	3 (3)
25	moe je anders doen	Consonant omission	moet je anders doen	4	5 (5)
26	ove de	Consonant omission	over de	2	2 (2)
27	in je leve wil	Consonant omission	in je leven wil	4	5 (5)
28	ik eb dat	Consonant omission	ik heb dat	3	3 (3)
29	to wel	Consonant omission	toch wel	2	2 (2)
30	zit no wel	Consonant omission	zit nog wel	3	3 (3)
41	geen moment was	Omission of 3 phonemes	gegeven moment was	3	4 (6)
42	ik kan volde keer	Omission of 3 phonemes	ik kan volgende keer	4	5 (6)
43	vos mij	Omission of 3 phonemes	volgens mij	2	2 (3)
47	of het mook is	Omission of 4 phonemes	of het mogelijk is	4	4 (6)
49	is oveer wat	Omission of 3 phonemes	is ongeveer wat	3	
50	ik weet eik ook	Omission of 5 phonemes	ik weet eigenlijk ook	4	

* Number of syllables of the stimulus followed by, in brackets, the number of syllables of the correct response