

# LiCoS: A New Linguistically Controlled Sentences Test to Assess Functional Hearing Performance

Martine Coene<sup>a, c</sup> Stefanie Krijger<sup>b</sup> Eline van Knijff<sup>a</sup> Matthias Meeuws<sup>c</sup>  
Geert De Ceulaer<sup>c</sup> Paul J. Govaerts<sup>a–c</sup>

<sup>a</sup>Language and Hearing Center Amsterdam, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands;

<sup>b</sup>Department of Otorhinolaryngology, Ghent University, Ghent University Hospital, Ghent, Belgium;

<sup>c</sup>The Eargroup, Antwerp, Belgium

## Keywords

Speech audiometry · Sentence understanding · Linguistic representation · Adult listeners

## Abstract

**Purpose:** To overcome the potential tension between clinical and ecological validity in speech audiometric assessment by creating a new set of sentence materials with high linguistic validity for the Dutch-speaking area. **Methods:** A linguistic “fingerprint” of modern spoken Dutch and Flemish served to generate a set of sentences recorded from 1 male and 1 female talker. The sentences were presented to 30 normal-hearing listeners in stationary speech noise at a signal-to-noise ratio (SNR) of –5 dB sound pressure level (SPL). A list design criterion was used to achieve perceptive homogeneity across the test lists, by scrambling lists of sentences of different syntactic types while controlling for linguistic complexity. The original set of test materials was narrowed down to 360 sentences, and list equivalency was evaluated at the

audiological and linguistic levels. A psychometric curve was generated with a resolution of 2 dB based on a second group of 60 young normal-hearing native speakers of Dutch and Flemish. **Results:** Sentence understanding showed an average repetition accuracy of 63.40% (SD 1.01) across the lists at an SNR of –5 dB SPL. No significant differences were found between the lists at the level of the individual listener. At the linguistic level, the sentence lists showed an equal distribution of phonological, morphological, and syntactic features. **Conclusion:** LiCoS combines the clinical benefit of acoustic control at the list level with the high ecological validity of linguistically representative test items. The new speech audiometric test is particularly appropriate to assess sentence understanding in individuals who would otherwise exhibit near-ceiling performance when tested with linguistically more simplified test stimuli. In combination with pure tone audiometric assessment, LiCoS provides valuable complementary information with respect to the functional hearing of patients.

© 2018 S. Karger AG, Basel

## Introduction

### *Speech Audiometry as a Measure of Functional Hearing Performance*

To ensure that those who need hearing devices obtain and use them to a maximum benefit, the development and constitution of adequate hearing assessment and rehabilitation programs should meet the challenges that current populations with a hearing impairment are facing in daily communication. As oral communication is a social skill that heavily relies on the ability to hear and understand speech, pure tone audiometric results are therefore ideally complemented by speech audiometric testing.

A critical analysis of existing speech audiometric assessment materials for the Dutch- and Flemish-speaking area [1] has revealed that although existing speech audiometric sentence tests are without any doubt useful components of the diagnostic audiological test battery (see, for instance, the “VU Amsterdam sentence test” [2, 3] and LIST [4]), they are not very representative of the speech that patients are confronted with in day-to-day listening situations.

As it has been shown that one of the main reasons for the nonuse or underuse of hearing aids is the user’s poor benefit and listening experience in real-life communication situations [5], there is an increasing interest in the ecological validity of hearing tests that can be used in clinical audiological practice.

### *Why Linguistic Balancing Matters*

An important feature of modern speech audiometric tests is that they seek to achieve a high comparability of sentence lists by controlling for acoustic information, i.e., test lists are compiled in such a way that in normal-hearing listeners, the variation between each of them in terms of speech reception thresholds (SRTs) – expressed as the hearing level at which a performance level of 50% correctly repeated speech items is reached – is only minimal.

Importantly, research on German sentence perception tests for speech audiometry has revealed that in addition to acoustic variation, SRTs are also sensitive to linguistic parameters. Variation in syntactic structure, word frequency, and sentence ambiguity have been shown to have a consistent and systematic effect on sentence intelligibility test outcomes. For sentences that are presented in noisy conditions, linguistic complexity has a negative effect on SRTs in young listeners with normal hearing [6] and on elderly listeners with hearing loss, especially in combination with age-related cognitive decline [7]. In a similar vein, noise has been shown to have a stronger im-

pact on the processing and understanding of structurally difficult parts of the sentence than on syntactically simpler parts [8].

### *Study Objectives*

Taken together, the combination of acoustic and linguistic features for speech perception may have important implications for speech audiometric assessment in audiological practice. Nonauditory factors such as lexical and syntactic features of the target language system may increase the cognitive demands of processing sentences in noise. In combination with hearing loss this may lead to suboptimal functional hearing in day-to-day listening situations even among patients with good speech discrimination outcomes. In this sense, the use of a balanced list of test sentences with varying degrees of syntactic complexity may provide useful information with regard to the subjective benefits of particular hearing devices for a patient.

The goal of the present study was to develop and evaluate a new set of sentence lists for the Dutch- and Flemish-speaking area that takes into account both acoustic and linguistic parameters that might affect speech audiometric outcomes. In addition, our intention was to develop materials that may be used to evaluate the hearing performance of hearing-impaired individuals in communicative settings that are representative of the modern Dutch language as it is used in daily communication. This implies that the sentences should be as close as possible to genuine conversational speech, and as such be uttered at a normal speaking rate. Against this background of ecological validity, linguistic control is ideally obtained by carefully balancing the lexical and grammatical complexity levels within each test list, reducing as much as possible any potentially unwanted linguistic effect on speech perception outcomes.

## Methods

### *Linguistic and Audiological Selection Criteria for Sentence Generation and Recording Principles*

During a first stage, a set of 500 sentences was created in such a way that they contain a fixed number of keywords (2 per sentence) on the basis of which sentence intelligibility was assessed. As semantic predictability is known to facilitate word recognition, all sentences were generated in such a way that they comply with the syntactic rules of modern Dutch but have very low semantic transparency, i.e., no fixed expressions have been used, nor do the 2 keywords within 1 sentence belong to the same semantic field.

Importantly, to achieve a representative yet homogeneous level of linguistic variation, each list was balanced for a large number

of linguistic parameters covering the lexical, phonological, morphological, and syntactic components of modern spoken Dutch: *lexical balancing* consisted in selecting the keywords from a list of the 5,000 most frequently used words in Dutch based on a large reference corpus (9 million words) of contemporary Dutch as spoken in the Netherlands and Flanders by adult speakers (Corpus Gesproken Nederlands [9]); *phonemic balancing* was obtained by calculating the Euclidean distance between the normalized histograms for the initial phonemes and the phoneme bigrams of the keywords and the abovementioned reference corpus of spoken Dutch [for a detailed description of the algorithm, see 10]; and the keywords were balanced for 8 major *word classes* (adjectives, adverbs, nouns, prepositions, pronouns, quantifiers, verbs, and conjunctions), for which the relative proportion reflects the distribution found in the Dutch reference corpus.

Importantly, all sentences were also balanced for *syntactic structure*, providing a representative sample of the variation in syntactic complexity that is found in modern spoken language. Each sentence list thus covers 8 different types, ranging from syntactically “simple” main clauses (e.g., with the canonical word order *Subject-Verb-[Object]*) over clauses with “medium” complexity (e.g., with noncanonical word orders such as the ones found in passives or topic-verb constructions) to “fully complex” structures (e.g., coordinated and subordinated clauses, subject and object relatives). Again, the relative proportion of each syntactic type is in agreement with the large reference corpus of modern spoken Dutch.

All lists were further balanced with respect to the *length of the sentences*, reducing variation within one syntactic type to maximally 1 word and to 2 syllables per sentence. In a similar vein, the *length of the keywords* was also balanced, resulting in limited variation in the proportion of mono-, bi-, tri-, and quadrisyllabic words per test list.

The 500 sentences were recorded by 1 male and 1 female speaker. Recordings were done separately for Dutch (The Netherlands) and Flemish (Flanders, Belgium) native speakers, using a H2 ZOOM recorder connected to an Acer laptop computer with a Realtek High Definition Audio sound card. Signals from the microphone were captured through the Praat software [11] at a sample frequency of 44,100 Hz. The total set of recordings thus consisted of  $2 \times 500$  sentences uttered by 2 male and 2 female speakers. The male and female voices were carefully balanced across the syntactic types. The 4 speakers were trained speech-language pathologists, and their speech was evaluated by 2 independent and naive observers who judged it not to present any regional pronunciation.

The average speaking rate of the sentences was 4.5 syllables per second, a rate which is in line with the references in the literature regarding normal conversational speech [12, 13]. All sentences were uttered with a deliberately flat intonation and with no pitch accents standing out on the keywords. They were saved as unique sound files in Waveform Audio File Format (.wav).

Further acoustic balancing consisted in adjusting the root mean square level of the individual sentences to 70 dB sound pressure level (SPL). A stationary speech noise was then created by shaping a white noise according to the long-term average spectrum of the complete set of sentences using a Praat script based on code developed by Quené and van Delft [14]. A post hoc analysis showed that in agreement with the principles laid out in the ISO 8253-3 norms concerning audiometric test methods, the speech level of a single test sentence did not deviate from the average level of all sentence test items by more than  $\pm 3$  dB.

#### *Sentence Selection and New List Formation*

All sentences were inserted in AŞE® 2012 audiometric assessment software [15]. All correct and erroneous repetitions of the keywords within each carrier sentence were scored directly in the AŞE® 2012 software program by the test administrator, a native speaker of Dutch. The sentences were presented in the abovementioned speech noise at a signal-to-noise ratio (SNR) of  $-5$  dB in the sound field from a single loudspeaker at  $0^\circ$  azimuth on the horizontal axis at 1 m distance from the listener. All participants were tested in the same quiet room under similar acoustic testing conditions. Prior to testing, the noise levels of the room were checked by means of an MPAA1 T.meter.

The complete set of 500 sentences was presented to 30 listeners (8 men and 22 women with an average age of 27.2 years [SD 9.9]). For each participant, the pure tone average (PTA) at 500, 1,000, 2,000, and 4,000 Hz was first estimated for the left and the right ear following the modified Hughson-Westlake procedure [16] using an Interacoustics AS608 type 4 audiometer and Amplivox Audio-cups headphones. At the group level, the average PTA was 11 dB HL (SD 2.24) for the left ear and 10.45 dB HL (SD 6.12) for the right ear. The 30 listeners were instructed to repeat what they had heard and were encouraged to guess if they had missed a word or part of a sentence. In total, the sentence repetition task thus yielded 15,000 sentence repetitions. For each sentence, a percent correct score was calculated based on  $30 \times 2$  keyword repetitions.

Based on these test results, within each syntactic type and for each speaking voice (male, female) the sentences were then sorted by percent score and outliers were removed from the test set. From the initial set of 500 sentences, 360 were retained and assigned to 12 lists containing 30 sentences each. Each list was carefully balanced based on the linguistic parameters described above, controlling for the length and syntactic structure of the carrier sentences and for the length of the keywords. In agreement with the methodology used for the development of the AzBio sentence lists for English [17], the first 12 sentences were sequentially assigned to lists 1–12 and the next 12 to lists 12–1. This was done for each syntactic type and speaking voice separately.

#### *Norm References*

Normative data were then gathered for the Flemish and Dutch listeners separately using the corresponding recordings of the sentence test. A test protocol was designed in view of generating a psychometric curve with a resolution of 2 dB based on speech repetition accuracy. The participants in this norm reference study were 60 young normal-hearing listeners who were native speakers of one of the regional variants of the Dutch language ( $n = 30$  Flemish;  $n = 30$  Dutch). The inclusion criteria for participation were based on the age (between 18 and 35 years old) and the hearing threshold of the listener (PTA  $\leq 20$  dB HL for both ears).

The speech audiometric assessment then started with the presentation of one of the 12 lists, for which the intensity of the speech signal was set at 78 or 76 dB SPL (randomly distributed amongst the participants) and against a background noise with a fixed intensity level of 70 dB SPL. Speech perception performance was measured in terms of the percentage of correct sentence repetitions based on 60 keywords per list of 30 sentences (i.e., a sentence repetition accuracy of 70% indicates that 42/60 keywords were repeated correctly). For the second and successive lists, the intensity of the speech signal was increased by 4 dB and testing was continued until a 100% repetition score was obtained (i.e., 60/60 correct

keyword repetitions in a list of 30 sentences). The initial level was then decreased by 4 dB until a speech repetition accuracy of 0% was reached (i.e., 0/60 correct keyword repetitions in a list of 30 sentences).

The second part of the speech audiometric assessment consisted in a presentation of one of the 12 sentence lists in silence at an intensity level of 40 or 42 dB SPL (randomly distributed amongst the participants) and followed by a stepwise increment of 4 dB for each consecutive list until a 100% repetition score was reached. The subsequent list was presented at a decreased intensity of 4 dB. Again, the initial level was then decreased by 4 dB until 0% repetition accuracy was obtained.

As such, for each listener, a set of different values of speech repetition accuracy between 0 and 100% was obtained for both listening conditions (silence, noise).

## Results

### *List Equivalency at the Level of Speech Repetition Accuracy*

A first analysis of sentence intelligibility based on the repetition accuracy showed an average correct score of 63.40% (SD 1.01) across the lists (see Table 1 for more details). The new test set of 360 sentences was further evaluated in view of the equivalency of the different lists both at the group level and at the level of the individual listeners. Firstly, the between-list variability at the group level was tested by running a one-way ANOVA with the percent correct scores for all 30 listeners per list as the dependent factor and the list number as an independent group factor. No significant difference was found between the 12 lists ( $F(11, 348) = 0.463, p = 0.969$ ), indicating an equivalency in perceptual difficulty for all lists. Secondly, a one-way within-subject ANOVA was conducted to assess the effect of list choice on the sentence repetition accuracy for each individual listener. Here also, no significant difference was found between the 12 lists ( $F(11, 348) = 0.892, p = 0.549$ ), implying that the different lists are of equal difficulty at the individual level as well.

### *List Equivalency at the Language Level*

The variability in speech materials between the 12 lists was further assessed for the different linguistic parameters that were used as additional design criteria to compile the sentence test set. Firstly, the phonemic histograms of the 12 lists indicated that the phonemic distribution was similar to that of the spoken language reference corpus. The LiCoS lists exhibit an average phoneme  $z$ -distance of 2.08 (SD 0.29) to the reference corpus; as such, its test items may be considered to be “best phonemic exemplars” of spoken Dutch. With respect to the variation in

**Table 1.** Average percent correct scores, SDs, and 95% confidence intervals for each test list

List No.	Average		95% confidence interval	
	% correct	SD	lower bound	upper bound
1	64.22	8.52	61.05	67.40
2	63.00	10.40	59.12	66.88
3	63.94	9.92	60.24	67.65
4	62.56	8.00	59.57	65.54
5	64.39	9.99	61.03	67.74
6	65.17	9.17	61.74	68.59
7	63.94	9.73	60.31	67.58
8	64.06	9.59	60.48	67.64
9	62.11	8.21	59.05	65.18
10	62.78	9.77	59.13	66.43
11	62.72	9.83	59.05	66.39
12	61.94	7.78	59.04	64.85

syntactic complexity, full list equivalency was achieved due to the fact that a fixed number of sentences of each syntactic type was included in each list in combination with a careful balancing between the male and female speaker’s voices and sentence length based on the number of words within each syntactic type. The sentence test set proved to be highly homogeneous, exhibiting a very stable average length of 2 syllables (SD 0.085) for the keywords and of 11.69 syllables (SD 2.12) for the carrier sentences across all lists. An overview of the linguistic parameters for each sentence list is given in Table 2.

### *Materials for Clinical Use*

In clinical practice, there may be circumstances in which audiologists wish to perform a speech audiometric assessment based on smaller numbers of test sentences (e.g., due to lack of time or manpower on the side of the audiologist, or a reduced attention span in the case of particular groups of listeners such as elderly adults). Therefore, the 360 sentences were assigned once again to 36 lists of 10 sentences according to their repetition accuracy and in agreement with the sequential assignment principle described above. For this shortened test version, equivalency was assessed by comparing the average level of performance scores obtained from the lists with 30 sentences to the lists with only 10 sentences. A within-subject analysis of the sentence repetition outcomes of the 30 listeners as a dependent variable and the length of the sentence lists (10 sentences vs. 30 sentences) and the list number as independent variables revealed that there was no significant

**Table 2.** Linguistic parameters at the sentence list level: distribution of the syntactic types, number of words, and number of sentences per syntactic type and of the speaker voice with a representative example for each syntactic type

Syntactic type	Words/ sentence	Male speaker, sentences/ list	Female speaker, sentences/ list	Total, sentences/ list	Example
Subj V Complement	7–8	3	3	6	<i>Haar partner is naar een ander stad vertrokken</i> “Her partner has left for another city”
SV <sub>trans</sub> O	7–8	5	5	10	<i>Hij isoleert zijn pand om geld te besparen</i> “He isolates his house to save money”
Topic Verb Subj	6–7	2	1	3	<i>Gisteren is de schommel kapot gegaan</i> “Yesterday the swing broke”
Passive	6–7	1	2	3	<i>Toetsen worden door de ouders ondertekend</i> “Tests must be signed by the parents”
Coordination	9–10	1	2	3	<i>We vonden het erg leuk en bleven dus langer</i> “We liked it a lot and thus stayed longer”
Subordination	9–10	2	1	3	<i>Hij dacht niet dat jij die tafel zou kopen</i> “He didn’t think that you would buy that table”
Subject relative	9–10	1	0	1	<i>De koe die daar in de stal staat is ziek</i> “The cow that is standing in the stable over there is ill”
Object relative	9–10	0	1	1	<i>De broek die ik aan had, was van mijn zus</i> “The trousers that I was wearing were my sister’s”
Total		15	15	30	

Each list thus contains 30 sentences, 15 uttered by a male speaker and another 15 by a female speaker.

difference in hearing performance based on the length of the test list ( $F(1, 29) = 0.141, p = 0.710$ ), the list number ( $F(11, 319) = 0.281, p = 0.989$ ), or a combination of the two (length  $\times$  list  $F(11, 319) = 0.340, p = 0.976$ ).

#### Norm Reference Curves and Speech Recognition Thresholds

For each participant, 6–8 data points between 0 and 100% were obtained both for the silent and the speech-in-noise listening condition. Together, the data points cover stimulus presentation levels with an intensity between 16 and 48 dB SPL (silence) and SNRs between –16 and +14 dB (noise).

In agreement with standard practice in norm referencing, percentile ranking scores were calculated for the different intensity levels and SNRs for Dutch and Flemish separately. The Appendix provides the full normative data description in terms of the 2.5th, the 50th, and the 97.5th percentile values for each listening condition and regional variant of the Dutch language. A sigmoid curve was fitted through these data points based on the logistic model function in (1):

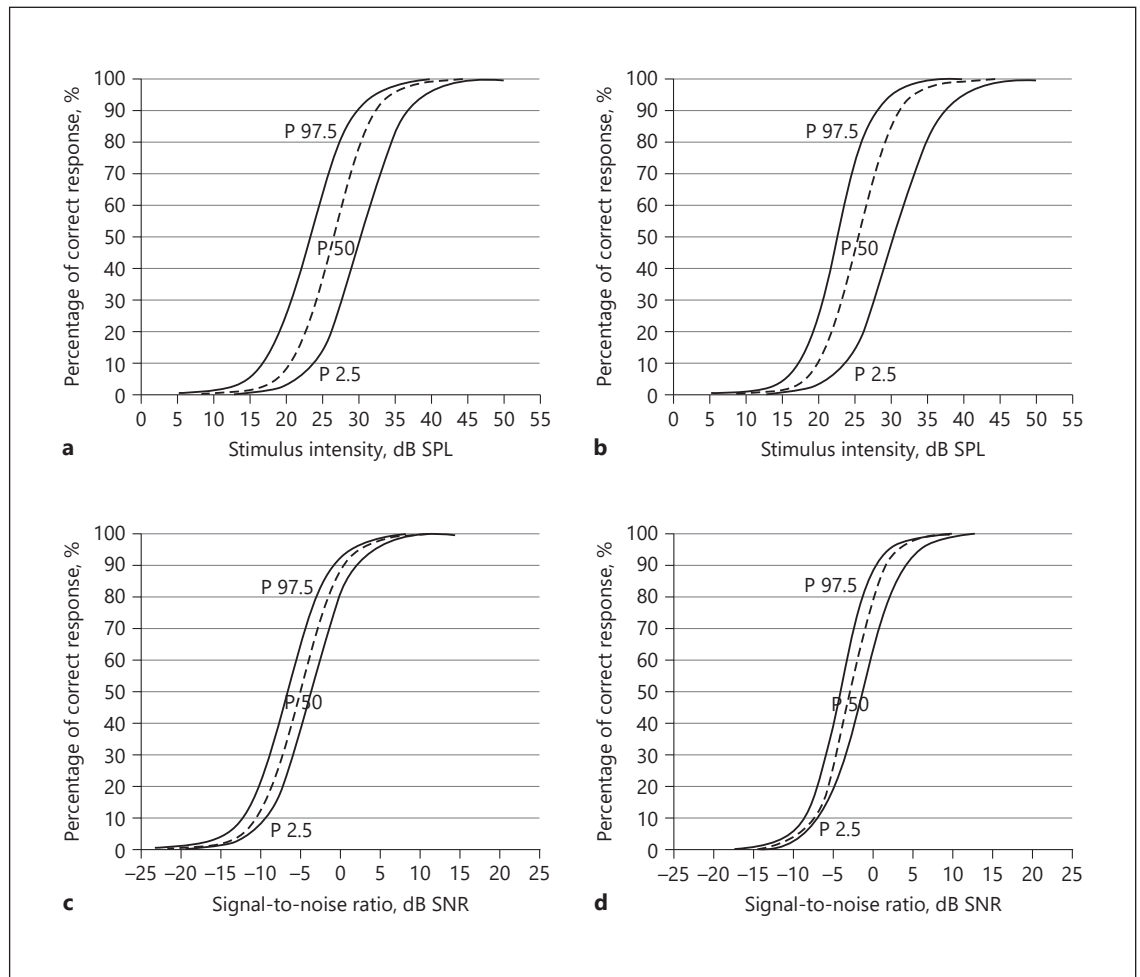
$$f(Int) = \frac{100}{1 + e^{-\alpha(Int - Int^o)}}, \quad (1)$$

where  $Int$  represents the intensity of the test stimulus,  $\alpha$  the steepness of the curve, and  $Int^o$  the intensity level at which the sigmoid function’s midpoint is reached (or the SRT at 50% correct repetitions). The percentile rankings are depicted in Figure 1. Details about the norm reference data can be found in the Appendix below.

In addition, for each listening condition and regional language variant, the average SRT and the SDs were calculated based on the individual speech identification curves obtained for each listener by applying the same logistic model function. The slope of the curve in its midpoint was calculated based on the function’s first derivative, for which the equation is given in (1’):

$$\frac{d}{dx} f(I^o) = f(I^o) \times (1 - f(I^o)) \times \alpha. \quad (1')$$

The resulting SRTs and slopes are given in Table 3 and the corresponding curves in Figure 2.



**Fig. 1.** Norm reference curves for sentence repetition accuracy depicting the 2.5th, 50th and 97.5th percentiles (P 2.5, 50, and 97.5). **a, b** Horizontal axes: speech stimulus intensity in quiet in dB sound pressure level (SPL). **c, d** Horizontal axes: intensity or signal-to-noise ratio (SNR) of the speech signal in dB SNR. **a-d** Vertical axes:

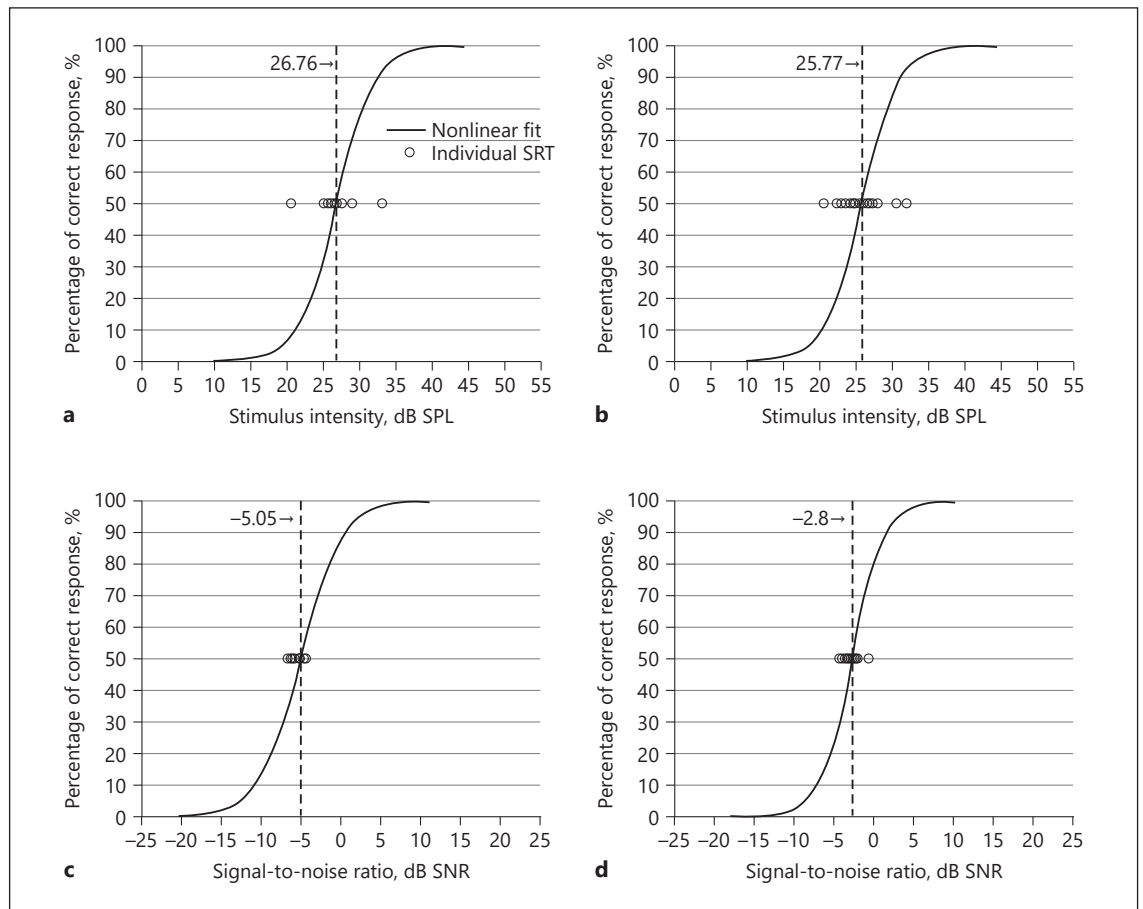
percentage of correctly repeated sentences based on 2 keywords per sentence. **a** Silence condition for Dutch listeners. **b** Silence condition for Flemish listeners. **c** Noise condition for Dutch listeners. **d** Noise condition for Flemish listeners.

## Discussion

Current sentence test materials for the Dutch-speaking area are typically well controlled for perceptive similarity. This is generally done by reducing linguistic variation up to the level of the individual test items, e.g., by using a particular sentence type exhibiting low syntactic complexity such as the *Subject-Verb-Object* template. This type of construction represents the canonical word order of Dutch with the agent and object of a transitive verb taking up, respectively, the sentence-initial and sentence-final positions. In such prototypical constructions, the listener may partially fill in missing acoustic informa-

**Table 3.** Speech reception thresholds (SRTs) and curve slopes at 50% for Dutch and Flemish listeners in silence and noise conditions

SRT	Dutch	Flemish
Silence	26.76 dB SPL (SD 2.28) slope at 50% = 9.9% (SD 1.6)	25.77 dB SPL (SD 2.21) slope at 50% = 10.20% (SD 2.31)
Noise	-5.05 (SD 0.72) slope at 50% = 9.94% (SD 1%)	-2.80 (SD 0.73) slope at 50% = 12.90% (SD 3.30)



**Fig. 2.** Speech reception thresholds (SRTs) for sentence understanding in silence (**a, b**) and noise (**c, d**) listening conditions for Dutch (**a, c**) and Flemish (**b, d**) listeners. The curve represents the average of the individual listener’s sentence repetition accuracy (average SRT based on individual nonlinear curve fitting; LiCoS). The dots at the curve’s midpoint depict the individual listener’s SRTs. **a, b** Horizontal axes: speech stimulus intensity in quiet in dB

sound pressure level (SPL). **c, d** Horizontal axes: signal-to-noise ratio (SNR) of the speech signal in dB SNR intensity or SNR of the speech signal in dB SNR. **a–d** Vertical axes: percentage of correctly repeated sentences. **a** Silence condition for Dutch listeners. **b** Silence condition for Flemish listeners. **c** Noise condition for Dutch listeners. **d** Noise condition for Flemish listeners.

tion based on syntactic predictability, e.g., by merely “guessing” the presence of the missed object of a verb based on the grammatical principle inherent in the Dutch language that transitive verbs do not allow their object to remain empty.

By ignoring such linguistic effects on sentence understanding, variation in speech audiometric outcomes may be wrongly attributed to sensory processing capacities only. Many clinics are using sentence intelligibility tests to evaluate the hearing performance of patients with a hearing impairment over time; when outcomes are based on linguistically simplified test sets, increased speech perception performance may be the result of good linguistic compensation skills, rather than genuine hearing capaci-

ties. This implies that such outcomes may potentially mask the need for further fine-tuning of hearing aids or cochlear implant speech processors. For the German-speaking area, these insights have led to the development of a new speech audiometric sentence test in which linguistic complexity is employed as an additional design criterion to compile sentence lists [18].

As is well known, natural conversation includes a wide variety of sentences, including complex syntactic structures that may be less supportive to speech perception and may even put a burden on auditory processing. Within our own group, we have recently reported important linguistic effects of test stimuli on hearing performance outcomes for the Dutch-speaking area [19]. Using test

sentences with varying degrees of linguistic difficulty at the morpholexical and syntactic levels, we have found that sentence repetition accuracy is significantly lower with verbs than with nouns, adjectives, or adverbs, or with syntactically more complex constructions such as passive sentences (e.g., *The document has been released into the public domain*) or object relative clauses (e.g., *I am looking for the book that I bought yesterday*) as compared to “easy” *Subject-Verb-Object* sentences.

Also, the LiCoS test sentences were deliberately recorded using a normal conversational style and speaking rate instead of a less naturalistic form of speech which is optimized for intelligibility (“hyperspeech”) [20, 21]. Due to the combination of the enhanced linguistic complexity and normal speech rate of the test items, the new LiCoS test provides a highly realistic impression of the differences in day-to-day hearing functioning in oral communication. As fewer subjects will achieve maximum performance in quiet or in noise conditions, these new LiCoS sentence test materials will be particularly suited for individuals who show near-ceiling performance when tested with linguistically less challenging items such as the ones used in the existing sentence tests [2, 4].

## Conclusions

The main aim of this study was to overcome the potential tension between clinical and ecological validity in speech audiometric assessment by creating a new set of sentence materials for the Dutch-speaking area (Flanders, The Netherlands) estimating a listener’s hearing performance in everyday listening situations. Within the limits of the ISO 8253-3 norm for speech audiometry [22], a design criterion has been used that aims to achieve homogeneity across test lists, by scrambling lists of sentences of different syntactic types while controlling for linguistic complexity at the list level. As such, in one speech audiometric test list, the clinical benefit of high acoustic-linguistic control is combined with the ecological validity of speech stimuli that are highly representative of modern spoken language.

A new sentence test set of 12 lists of 30 sentences was recorded in two standard varieties of the Dutch language spoken in the Netherlands and in Belgium (Flanders). The evaluation of the test materials in a cohort of 30 normal-hearing listeners indicated that LiCoS may serve as a valid measure of functional hearing performance. Each list is balanced for a large number of linguistic factors that cover the lexical, phonological, morphological, and syn-

tactic components of modern spoken Dutch while yielding a similar speech repetition accuracy under identical listening conditions.

Norm reference data were gathered based on another cohort of 60 normal-hearing listeners from the Netherlands and Belgium. These norm references may serve as a baseline to evaluate hearing performance in patients with a hearing impairment. Finally, the existence of a “shortened” version of LiCoS paves the way to the development of a Minimal Speech Test battery for Dutch, following the example of Spahr et al. [17] for English. Ideally, such a Minimal Speech Test battery would include a phonemic discrimination task [15], a list of CVC words [23], a digits-in-noise test [24], and 10 LiCoS sentences, providing an optimum mix of low-redundancy and highly redundant speech samples that could be used successfully to evaluate speech understanding of hearing-impaired patients in clinical audiological practice.

## Acknowledgements

The authors wish to thank Renske Berkmoes for her help in collecting the data from Dutch native listeners as part of an MA research project, jointly supervised at Vrije Universiteit Amsterdam by M. Coene and Paul J. Govaerts.

## Statement of Ethics

Ethical approval for this study was granted by the Ethics Committee of the Faculty of Humanities at Vrije Universiteit Amsterdam (VU EC02.14). In agreement with the principles expressed in the Declaration of Helsinki, participation was voluntary, and there were no rewards, costs, or risks attached to participating. All participants provided informed consent prior to partaking in the study.

## Disclosure Statement

The authors declare to have no conflicts of interest.

## Funding Sources

Our research has received funding from the Netherlands Organisation for Scientific Research, within the NWO Alfa Meerwaarde funding scheme (M. Coene); a PhD fellowship of the Research Foundation Flanders, FWO Vlaanderen, Belgium (S. Krijger); the Netherlands Organisation for Scientific Research (E. van Knijff); and the European Union’s 7th Framework Programme for Research, Technological Development and Demonstration under the FP7-PEOPLE-2012-IAPP project “Hearing Minds,” Grant Agreement No. 324401 (Paul J. Govaerts).



## Norm Reference Data for LiCoS in Silence and in Noise for Dutch and Flemish

Percentiles Speech in Silence Dutch															
dB SPL	15	17	19	21	23	25	27	29	31	33	35	37			
min	0.0	0.0	0.0	0.0	5.0	18.0	5.0	52.0	37.0	80.0	67.0	98.0			
P2.5	0.1	0.1	0.4	0.0	6.4	19.4	14.5	53.0	44.2	81.1	69.5	98.1			
P16	0.5	0.8	2.1	0.0	13.7	27.0	47.4	57.0	75.0	86.3	83.0	98.3			
P50	1.5	2.5	4.0	7.0	22.0	40.0	52.0	70.0	83.0	95.0	95.0	99.0			
P84	2.5	4.2	13.8	14.4	35.2	44.2	67.3	72.4	94.4	95.0	95.0	99.7			
P97.5	2.9	4.9	27.4	15.0	70.4	46.6	88.5	73.0	99.0	97.3	95.0	100.0			
max	3.0	5.0	30.0	15.0	78.0	47.0	93.0	73.0	100.0	98.0	95.0	100.0			

Percentiles Speech in Noise Dutch															
SNR	-16	-14	-12	-10	-8	-6	-4	-2	0	2	4	6	8		
min	0.0	0.0	0.0	0.0	17.0	30.0	43.0	60.0	85.0	85.0	92.0	97.0	97.0		
P2.5	0.0	0.0	0.0	2.5	17.4	32.8	44.8	61.8	85.0	86.8	92.4	97.0	97.1		
P16	0.0	0.0	2.0	8.5	18.5	38.0	49.2	67.7	87.2	92.0	95.5	97.0	97.5		
P50	0.0	0.0	3.0	13.0	23.0	43.0	58.0	75.0	90.0	93.0	97.0	97.0	98.5		
P84	0.0	3.5	7.8	17.8	31.5	49.5	65.0	84.5	93.0	96.5	99.5	100.0	99.5		
P97.5	0.0	5.0	9.3	20.6	37.6	55.9	67.0	85.0	94.3	97.7	100.0	100.0	99.9		
max	0.0	5.0	10.0	22.0	40.0	58.0	68.0	85.0	95.0	98.0	100.0	100.0	100.0		

Percentiles Speech in Silence Flemish																	
dB SPL	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48
min	0.0	0.0	0.0	0.0	0.0	12.0	28.0	55.0	53.0	66.0	79.0	88.0	85.0	90.0	98.0	100.0	100.0
P2.5	0.0	0.0	0.0	2.6	8.8	15.0	36.8	58.0	61.8	72.0	82.9	89.5	87.5	92.4	98.2	100.0	100.0
P16	0.0	0.0	1.5	8.8	31.6	35.0	56.0	70.0	82.1	87.4	90.4	92.4	95.2	99.8	100.0	100.0	100.0
P50	0.0	0.0	7.0	19.0	42.5	52.5	77.0	83.0	92.0	92.5	97.0	97.5	100.0	100.0	100.0	100.0	100.0
P84	1.6	5.0	24.6	34.6	60.2	71.1	85.2	93.0	95.2	97.0	98.2	99.5	100.0	100.0	100.0	100.0	100.0
P97.5	2.8	8.5	27.0	52.1	70.8	75.0	89.5	97.0	97.7	99.3	100.0	100.0	100.0	100.0	100.0	100.0	100.0
max	3.0	10.0	27.0	57.0	72.0	75.0	90.0	98.0	98.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Percentiles Speech in Noise Flemish															
SNR	-12	-10	-8	-6	-4	-2	0	2	4	6	8	10	12		
min	0.0	0.0	0.0	0.0	28.0	52.0	68.0	63.0	90.0	92.0	95.0	95.0	100.0		
P2.5	0.0	0.0	0.0	0.0	28.7	52.4	68.7	69.0	90.4	92.4	95.7	95.0	100.0		
P16	0.0	0.0	0.5	0.0	30.7	53.0	70.5	82.8	92.0	93.0	97.2	98.0	100.0		
P50	0.0	0.0	7.0	15.0	38.0	57.0	82.0	90.0	95.0	98.0	100.0	100.0	100.0		
P84	0.1	2.1	8.0	25.4	46.2	65.9	83.0	91.3	97.0	100.0	100.0	100.0	100.0		
P97.5	1.7	5.7	15.1	26.6	56.8	72.3	86.3	96.7	97.6	100.0	100.0	100.0	100.0		
max	2.0	7.0	17.0	27.0	62.0	74.0	87.0	98.0	98.0	100.0	100.0	100.0	100.0		

## References

- 1 Hammer A, Coene M, Govaerts P: Zinnen of woorden? Een bespreking van het spraakmateriaal binnen de Nederlandse en Vlaamse spraakaudiometrie. *Tijdschr Stem Taal Spraakpathol* 2013;18:1–13.
- 2 Plomp R, Mimpen AR: Improving the reliability for testing the speech reception threshold for sentences. *Audiology* 1979;18:43–52.
- 3 Versfeld NJ, Daalder L, Festen JM, Houtgast T: Method for the selection of sentence materials for the efficient measurement of the speech reception threshold. *J Acoust Soc Am* 2000;107:1671–1684.
- 4 van Wieringen A, Wouters J: LIST and LINT: sentences and numbers for quantifying speech understanding in severely impaired listeners for Flanders and the Netherlands. *Int J Audiol* 2008;47:348–355.
- 5 Kochkin S, MarkeTrak V: Why my hearing aids are in the drawer: the consumer's perspective. *Hear J* 2000;53:34–42.
- 6 Uslar V, Ruigendijk E, Hamann C, Brand T, Kollmeier B: How does linguistic complexity influence intelligibility in a German audiometric sentence intelligibility test? *Int J Audiol* 2011;50:621–631.
- 7 van Knijff EC, Coene M, Govaerts PJ: Speech understanding in noise in elderly adults: the effect of inhibitory control and syntactic complexity. *Int J Lang Commun Disord* 2018;53:628–642.
- 8 Carroll R, Ruigendijk E: The effects of syntactic complexity on processing sentences in noise. *J Psycholinguist Res* 2013;42:139–159.
- 9 Oostdijk N: Het Corpus Gesproken Nederlands: veelzijdig onderzoeksinstrument voor o.a. taalkundig en taal- en spraaktechnologisch onderzoek. *Link* 2003;14:3–6.
- 10 Hammer A, Vaerenberg B, Kowalczyk W, ten Bosch LF, Coene M, Govaerts PJ: Balancing word lists in speech audiometry through large spoken language corpora. *Proc Interspeech*. Lyon, 2013, p 3613.
- 11 Boersma P, Weenink D: Praat, a system for doing phonetics by computer. *Glott Int* 2001;5:341–345.
- 12 Wingfield A, McCoy SL, Peelle JE, Tun PA, Cox LC: Effects of adult aging and hearing loss on comprehension of rapid speech varying in syntactic complexity. *J Am Acad Audiol* 2006;17:487–497.
- 13 Quené H: Andante of allegro? Verschillen in spreektempo tussen Vlamingen en Nederlanders. *Onze Taal* 2008;77:179–181.
- 14 Quené H, van Delft LE: Non-native durational patterns decrease speech intelligibility. *Speech Commun* 2010;52:911–918.
- 15 Govaerts PJ, Daemers K, Yperman M, De Beukelaer C, De Saegher G, De Ceulaer G: Auditory speech sounds evaluation (ASE®): a new test to assess detection, discrimination and identification in hearing impairment. *Cochlear Implants Int* 2006;7:92–106.
- 16 Carhart R, Jerger JF: Preferred methods for clinical determination of pure-tone thresholds. *J Speech Hear Disord* 1959;24:330–345.
- 17 Spahr AJ, Dorman MF, Litvak LM, Van Wie S, Gifford RH, Loizou PC, Loisel LM, Oakes T, Cooks S: Development and validation of the AzBio sentence lists. *Ear Hear* 2012;33:112–117.
- 18 Uslar VN, Carroll R, Hanke M, Hamann C, Ruigendijk E, Brand T, Kollmeier B: Development and evaluation of a linguistically and audiolinguistically controlled sentence intelligibility test. *J Acoust Soc Am* 2013;134:3039–3056.
- 19 Coene M, Krijger S, Meeuws M, De Ceulaer G, Govaerts PJ: Linguistic factors influencing speech audiometric assessment. *Biomed Res Int* 2016;2016:7249848.
- 20 Smiljanić R, Bradlow AR: Speaking and hearing clearly: talker and listener factors in speaking style changes. *Lang Linguist Compass* 2009;3:236–264.
- 21 Cooke M, King S, Garnier M, Aubanel V: The listening talker: a review of human and algorithmic context-induced modifications of speech. *Comput Speech Lang* 2014;28:543–571.
- 22 International Organization for Standardization: ISO 8253-3:2012: Acoustics – Audiometric Test Methods – Part 3: Speech Audiometry. <https://www.iso.org/standard/45101.html>.
- 23 Bosman AJ, Wouters J, Damman W: Realisatie van een cd voor spraakaudiometrie in Vlaanderen. *Logop Audiol* 1995;9:218–225.
- 24 Smits C, Goverts ST, Festen JM: The digits-in-noise test: assessing auditory speech recognition abilities in noise. *J Acoust Soc Am* 2013;133:1693–1706.