



**HAL**  
open science

## Criteria for creating new standard reading passages for the assessment of speech and voice: A Delphi consensus study

Timothy Pommée, Mathieu Balaguer, Julie Mauclair, Julien Pinquier,  
Virginie Woisard

### ► To cite this version:

Timothy Pommée, Mathieu Balaguer, Julie Mauclair, Julien Pinquier, Virginie Woisard. Criteria for creating new standard reading passages for the assessment of speech and voice: A Delphi consensus study. *Clinical Linguistics & Phonetics*, 2022, 37 (8), pp.722-741. 10.1080/02699206.2022.2080589 . hal-04595275

**HAL Id: hal-04595275**

**<https://hal.science/hal-04595275>**

Submitted on 31 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Criteria for creating new standard reading passages for the assessment of speech and voice: A Delphi consensus study**

Timothy Pommée<sup>1\*</sup>, Mathieu Balaguer<sup>1,2</sup>, Julie Mauclair<sup>1</sup>, Julien Pinquier<sup>1</sup>,  
Virginie Woisard<sup>2,3,4</sup>

<sup>1</sup> *IRIT, CNRS, Paul Sabatier University Toulouse III, Toulouse, France*

<sup>2</sup> *ENT department, University Hospital of Toulouse Larrey, Toulouse, France*

<sup>3</sup> *Oncorehabilitation unit, University Cancer Institute of Toulouse Oncopole, Toulouse, France*

<sup>4</sup> *Laboratoire Octogone Lordat, Jean Jaurès University Toulouse II, Toulouse, France*

## **Correspondence to:**

\*Timothy Pommée

timothy.pommee@irit.fr

+33 6 03 03 51 48

ORCID: 0000-0001-7846-7282

IRIT Institut de Recherche en Informatique de Toulouse

118 Route de Narbonne

31062 TOULOUSE CEDEX 9

# **Criteria for creating new standard reading passages for the assessment of speech and voice: A Delphi consensus study**

Standard reading passages allow for the study of the integrated functions of speech and voice components in contextual, running speech, with target stimuli in a controlled environment. In both clinical practice and research, these texts provide rapid insight into the characteristics of the patient's speech, with fewer hesitations than in conversational speech and better predictability by the evaluator. Although a plethora of texts exist in different languages, they present various limitations. A specifically created standardized text in each language allowing for an ecological assessment of speech and voice functions, meeting most required criteria for standard speech and voice assessment and adapted to the target language's cultural and linguistic specificities, would therefore be an interesting option. However, no guidelines exist for the creation of such a reading passage. This article describes the international Delphi consensus study carried out to identify a minimal set of criteria to take into account when creating standard reading passages for an overall speech and voice assessment in adolescents and adults. This survey was conducted in three consecutive rounds; forty experts participated in the first round, with a total dropout of 17% from round 1 to round 3. It results in a minimal set of ten criteria which were selected by a majority of the experts and were rated as most important. This set contains five phoneme-level, two word-level, two sentence-level criteria and one global-level criterion. It can be used as a general guideline for the creation of standard reading passages in Indo-European Romance and Germanic languages such as English, French and German. The construction of a new reading passage in French following this guideline is briefly described.

Keywords: reading passage, text, speech, voice, assessment, Delphi

## **Introduction**

Speech can be assessed perceptually or instrumentally on samples of different levels of granularity (e.g. single words vs sentences) which are more or less functional (e.g., reading aloud from a text vs conversational speech sample) (Pommée, Balaguer, Mauclair, et al., 2021b). A wide variety of assessment tasks and tools exist, both in clinical practice (Gurevich

& Scamihorn, 2017; Pommée, Balaguer, Mauclair, et al., 2021a) and in scientific research (Pommée, Balaguer, Piquier, et al., 2021).

Each type of unit allows for an assessment that serves a specific purpose, which may differ depending on whether the assessment is perceptual or instrumental. In this article, we will focus more specifically on texts. Indeed, while segmental units represent the main unit of analytical intelligibility assessment (Fredouille et al., 2019; Pommée, Balaguer, Mauclair, et al., 2021b; Woisard et al., 2013), the use of a reading passage allows the speech assessment to be integrated into a more natural production context for a more functional assessment.

Examples of speech and voice assessment types that can be carried out on reading passages are pausing and breathing patterns, voice quality, pitch and intensity, disfluencies, the use of prosodic patterns, vowel and consonant articulation accuracy and velopharyngeal function. Thus, standard reading passages are of major interest in patient populations with sufficient reading proficiency, because they allow for the study of the integrated functions of speech and voice components in contextual, running speech (closest to spontaneous speech as compared to isolated words or sentences), with target stimuli in a controlled environment (Patel et al., 2013). In both clinical practice and research, these texts provide rapid insight into the characteristics of the patient's speech (Auzou & Rolland-Monnoury, 2006), with fewer hesitations than in conversational speech (Vasilescu et al., 2004) and better predictability by the evaluator.

However, although a plethora of texts exist in different languages, none of them seem to really meet the expectations and needs of clinicians and researchers for routine speech and voice assessment. Some were created for a specific purpose, such as the Zoo Passage in English (Fletcher, 1972) for the assessment of velopharyngeal closure, or the texts by Kuo and Weismer (Kuo & Weismer, 2016) for the assessment of vowel reduction. For most of the

others, little information is available about how they were created. Some are simply excerpts from literary works (e.g., Alphonse Daudet's 'La chèvre de Monsieur Seguin' (Daudet, 1869) in French), or translations of texts such as "The Northwind and the Sun" (International Phonetic Association, 1999) without adaptation to the target language (Jesus et al., 2015). A standardized text in each language allowing for an ecological assessment of speech and voice functions, meeting most required criteria for standard speech and voice assessment and adapted to the target language's cultural and linguistic specificities, would therefore be an interesting option. Specific speech stimuli, e.g. for the assessment of velopharyngeal function, could then be used once the main symptoms have been identified. However, we noticed the absence of guidelines for the creation of such a reading passage, which would benefit clinicians and researchers from various linguistic communities. Hence, the aim of this article is to describe the international Delphi consensus study we carried out to identify a minimal set of criteria to take into account when creating standard reading passages for the assessment of speech and voice functions. The reader should be aware that this study targets reading passages for the assessment of articulation (dysarthria, apraxia), prosodic variations and phonatory behavior (e.g. dysphonia, vocal feminization) and fluency disorders (stuttering/stammering) in patients whose reading fluency is sufficiently developed. It is therefore not adapted for developmental speech disorder assessment in children, nor intended for the assessment of reading proficiency or for language assessment (e.g. aphasia).

In this article, we will describe the selection and prioritization of the construction criteria by the international expert group.

## **Materials and methods**

### ***Participants***

We targeted professionals (clinicians, researchers, lecturers) engaged in activities in speech sound disorders<sup>1</sup> and/or fluency disorders (stuttering/stammering). ‘Activities’ were defined as clinical activity, research, academic or industrial activity, or a combination of these, if at least approximately 20% related to speech (self-estimated by the participants).

Recruitment was carried out via:

- national professional associations (speech-language-hearing, phoniatrics, voice, acoustics, computer science/signal processing, linguistics/phonetics); over 200 organizations were contacted worldwide by e-mail;
- social networks (Twitter, Facebook), where targeted professions and associations were also solicited in private groups;
- e-mail to over 50 hand-selected speech experts identified in literature searches on PubMed, who had at least three publications in the field of speech, were authors of a reference book, or participated in research projects linked with pathological speech.

Non-respondents for each round were excluded from subsequent rounds. An *a posteriori* statistical comparison using McNemar’s Chi<sup>2</sup> for paired samples was carried out to verify that the expert panel characteristics did not significantly change with the dropouts. It was also verified that the increasing consensus throughout the Delphi rounds was not biased by the dropouts of predominantly disagreeing experts. To that end, quantitative data in all three rounds with and without the dropouts were compared.

Forty experts completed round 1; thirty-four experts completed round 2 (85%); thirty-three experts completed round 3 (97%). A total dropout of 17% from round 1 to round 3 was observed.

---

<sup>1</sup>See the ASHA’s definition of speech sound disorders: <https://www.asha.org/practice-portal/clinical-topics/articulation-and-phonology/>

Detailed data for the participants in each round are available in Appendix A. The trends described hereafter are constant throughout the three rounds despite the dropouts. No statistical difference was found between the expert characteristics with and without the dropouts. A majority of the expert panel were speech and language pathologists (SLPs) working in the fields of speech, fluency and voice disorders. Other major groups of participants were linguists, ENT/phoniatricians and computer scientists. More than half of the experts had at least 10 years of experience working in the speech and voice domains. Their main activity was research for about 40%, clinical practice and academic activity for about a third of the experts, respectively. The vast majority of experts were engaged in at least two main activities; clinical activity and research were combined in half of the cases. More than half of the participants had a third-cycle diploma (PhD) obtained on average in 2009–2010 ( $\pm 8$  years).

France, the United Kingdom and Germany were the most represented countries, while the most frequent main language spoken at work was English, followed by French and German. The patient/study populations showed a slight prevalence for the elderly population. Also, the most encountered were acquired and degenerative neurological pathologies.

### ***Ethics approval***

This research was registered with the data protection officer of the *Centre National de la Recherche Scientifique* (CNRS) and was also approved by the computer science ethics advisory board (*Comité consultatif d'Éthique concernant la Recherche en Informatique de Toulouse* — CLERIT). An information sheet describing the purpose of the study, the participant's rights and the data privacy policy was provided to all participants prior to the first round of the survey. Participants gave consent for their answers to be used in an

anonymized and aggregated manner to derive consensus statements. Only e-mail addresses were known to the moderator, in order to enable round-to-round survey monitoring.

### *Study design*

In general, the Delphi methodology is conducted over several consecutive rounds (usually three) (Birko et al., 2015; Linstone & Turoff, 2002). In a ‘modified Delphi study’, statements to be rated are directly provided in the first round, based on a preliminary literature search (Cunningham et al., 2019; Denman et al., 2019). After each round, the panel’s responses are synthesized, and areas of agreement/disagreement are identified. Aggregated controlled feedback (von der Gracht, 2012) is then provided to the panel in the following round to explain modifications that have been made to facilitate consensus, and participants give their opinion on the new assertions (Denman et al., 2019; Diamond et al., 2014). The iterations are usually carried out until consensus is reached or until stability in the answers is observed (see (Chalmers & Armour, 2019; von der Gracht, 2012) for examples of stability measures).

The data discussed in the present article are part of a broader modified Delphi study which was designed in two parts. The first part, addressing the definition and measures of intelligibility and comprehensibility, is described in another article (Pommée, Balaguer, Mauclair, et al., 2021b). The present article addresses the second part, i.e. about the criteria for creating standard reading passages. The expert panel is thus the same for both studies. The stages of this second part of our Delphi study are summarized in Figure 1. The main steps which will be described hereafter are the problem definition, the literature search, the panel selection and the construction of the three consecutive Delphi rounds.

### **[FIGURE 1 NEAR HERE]**

#### *1. Definition of the problem*



As explained in the introduction, most available reading passages are either excerpts from literature works or translations, or were created to target a specific speech measure. To allow for the creation of new standard reading passages for the assessment of speech and voice, guidelines are necessary. Therefore, the aim of this Delphi study was to draw up a consensual minimal set of criteria for the creation of such reading passages.

## *2. Identification of criteria*

Prior to launching the survey, a targeted non-systematic literature search was carried out to identify commonly used reading passages in English, French and Dutch, as well as the criteria that were used to construct these passages. This search was carried out using PubMed as well as reference lists of known articles on the topic.

A non-extensive list of reading passages in French, English and Dutch, as well as their reported characteristics and shortcomings, can be found in Appendix B.

Four articles describing reading passage creations in detail were retrieved and were used as the primary source for developing an initial list of criteria:

- Patel et al., 2013: their goal was to provide a new English text that was better suited for speech assessment and differential diagnoses compared to the ‘My Grandfather’ reading passage (Van Riper, 1963);
- Martens & Nuffelen, 2010: these authors developed a Flemish text as part of a new protocol for assessing the prosodic skills of dysarthric patients;
- Powell, 2006: in this study, the phonetic, lexical and structural characteristics of 15 English texts were analysed and compared;
- Kuo & Weismer, 2016: their aim was to investigate intra- and inter-individual variations in vowel articulation in healthy speakers of American English.

The criteria identified in these articles were then enriched by an exchange with the authors of the French MonPaGe protocol (Laganaro et al., 2021), who had written a text specifically created for the assessment of dysarthric and apraxic speech and validated on French-speaking talkers from Belgium, France, Quebec and Switzerland (Fougeron et al., 2019; Pernon et al., 2020).

Published in 2021, it is the only French text to date for which the construction criteria have been detailed by the authors. Indeed, they constructed a short text (fewer than 200 words) that includes:

- words also presented in isolation in the protocol (e.g., days of the week also found in an automatic speech task of the protocol);
- glides;
- word repetitions;
- elements that allow for the assessment of vowel-to-vowel coarticulation (/papa/ vs. /papi/);
- the cardinal vowels /a,i,u/ in monosyllabic consonant-vowel (CV), CVC, and bisyllabic CVCV words;
- the same sequence at the beginning and end of the text (to assess fatigability);
- dialogues and melodic variations;
- predominantly nasal segments that allow for the evaluation of nasality;
- complex syllabic sequences.

However, these criteria were not based on an inventory of the needs of clinicians and researchers.

Based on this collaboration, as well as on the literature search, the following list of 25 potential criteria to consider when creating a new reading passage was established. The rationale for these 25 criteria can be found in Appendix C.

*Phoneme-level criteria:*

- (1) Complete phonemic inventory (i.e. the text includes all of the target language's phonemes at least once);
- (2) Phonetic balance (i.e. the included phonemes occur at approximately the same frequency at which they occur in ordinary conversation in the target language);
- (3) Inclusion of bi- and triconsonantal clusters;
- (4) Taking into account the positions of the phonemes in the words (beginning/middle/end), as well as their phonemic context;
- (5) Repeated inclusion of cardinal vowels;
- (6) Taking into account vowel-to-vowel coarticulation (e.g. word pair /papa/-/papi/);
- (7) Inclusion of initial voiceless stops and clusters, as well as unvoiced-voiced phoneme sequences (fluency disorders);
- (8) Occurrence of words with glides (lingual articulatory dynamics, dysarthria) and words inducing lingual movements across the vowel triangle (e.g. /kajak/, /kiwi/);

*Word-level criteria:*

- (9) Control of the ratio of function/content words (fluency disorders);
- (10) Control of the articulatory/phonetic complexity of words;
- (11) Repeated occurrence of some words (error consistency, apraxia vs dysarthria);
- (12) Inclusion of word pairs of increasing length (e.g. amuse/amusement);
- (13) Control of lexical frequency;
- (14) Coupling with items from isolated assessment tasks to compare isolated vs continuous speech performance (e.g. include words from single-word lists of assessment batteries);
- (15) Control of the phonological neighbourhood of words;

*Sentence-level criteria:*

- (16) Presence of various intonation patterns (questions, statements, contrastive emphasis...);
- (17) Inclusion of predominantly nasal vs. oral sentences (velar insufficiency, cleft palate...);
- (18) Presence of a 100% voiced sentence;
- (19) Repetition of an identical segment at the beginning and at the end of the passage (fatigue effect);
- (20) Inclusion of various sentence lengths (breathing support, fatigability...);
- (21) Control of the mean length of utterance;
- (22) Control of the syntactical complexity;

*Global-level criteria:*

- (23) Calculation of a readability index (e.g., Flesch-Kincaid);
- (24) Control of the topic (e.g., 'modern' topic, emotionally neutral content to avoid emotional charging and minimize memory effects...);
- (25) Control of the overall length of the reading passage.

### *3. Delphi rounds and data analysis*

This Delphi consensus survey was conducted in three consecutive rounds, between July and December 2020. The termination criterion defined prior to data analysis was either consensus on all items – i.e., agreement by at least 75% of the expert panel (Denman et al., 2019; Diamond et al., 2014) – , or stability of responses.

Round 1 was available for two months and a half; rounds 2 and 3 were available for one month. The first questionnaire was open access, with built-in duplicate checking and security procedures. The subsequent questionnaires were restricted to previous participants and required an individual single-use ID provided by the moderator. All questionnaires were

piloted by five researchers to get an estimate of the response time and to detect and correct potential execution problems (glitches and logical structure issues).

In each round, participants who did not agree with a statement were encouraged to explain the reason of their disagreement, but comments were not mandatory so as not to bias towards positive answers.

The qualitative and quantitative data were analyzed using descriptive statistics (frequencies, means, standard deviations, medians, interquartile ranges) on Stata/MP software (version 14, StataCorp, College Station, TX).

*Round 1.* The first round was constructed in two parts.

In the first part, participants listed the measures or ratings they wished to be able to use on a standardized reading passage in open-ended responses (at least 3). Two raters, blinded to the identity of the participants, carried out conventional content analysis (Denman et al., 2019; Hsieh & Shannon, 2005) of these open-ended responses to identify semantic groupings.

The second part addressed the selection and prioritization of the most important criteria for the construction of a standard reading passage. Participants were asked to select and rank, among the list of 25 criteria described above (see “Study design – 2. Identification of criteria), those they considered essential to control for the creation of a new reading passage. Any criteria that were not selected were considered non-essential; additional criteria could be suggested by the participants, specifying the target language. The responses and comments from this first round were analysed independently and anonymously by the same two raters.

Three measures were made for each criterion: mean rank (MR), the frequency with which a criterion was rated (and thus was considered by the participant to be essential; rating frequency, RF), and the mean rank when rated (MRR, thus excluding non-ratings).

For the MRs, non-ranked criteria were assigned the lowest value of 26 (because there were 25 available criteria). This measure is a composite indicator, taking into account not only the number of times each criterion was rated, but also the rank assigned to it at each rating.

RFs and MRRs, on the other hand, allow for different nuances of information: while the MRR informs us about the relative importance of each criterion for the group of experts who selected it, the RF informs us about the proportion of experts who judged a criterion as relevant, without, however, taking into account the relative importance assigned to it. The thresholds for a criterion to be considered essential to the panel of experts were set *a priori*: a criterion had to be ranked by at least half of the participants (i.e., the majority of experts found it important), and its MRR had to be less than  $7.5/25$  (33%) (i.e., it had to be ranked important by those who selected it).

*Round 2.* Responses and comments from round 1 were analysed in an anonymized manner by the same two raters. They were then synthesized and fed back to the participants for contextualization in round 2 (i.e. reminder of the task used in round 1, followed by a summarized overview of the quantitative and qualitative results), together with four new statements based on the previous response.

In this second round, participants rated a set of the most and best ranked criteria from the first round, as well as three additional statements regarding further criteria (inclusion of pseudowords based on participants' comments, as well as two mixed-opinion criteria from

round 1). All statements were rated using binary answers (Agree/Disagree), with optional comments.

*Round 3.* A third round was necessary to clarify one statement, which was found to be somewhat ambiguous in round 2.

## **Results**

### ***Round 1***

#### *Measures and ratings*

Table 1 summarizes the measures and ratings participants want to be able to use on a standard reading passage according to their open-ended responses (by decreasing order of mentions).

**[TABLE 1 NEAR HERE]**

#### *Selection and prioritization of criteria*

Figure 2 shows the 25 presented criteria and their average ranking by the participants. The eight most and best ranked criteria are listed in Table 2, in decreasing order of the mean rank (MR, out of 25).

**[TABLE 2 NEAR HERE]**

The two most poorly ranked criteria also had the smallest occurrences of selection:

- Syntactical complexity: 23.5 (20%; MRR: 13.3)
- Presence of a 100% voiced sentence: 23.3 (18%; MRR: 10.6)

Two other criteria were highlighted in the analysis, in light of the adverse results between the number of occurrences and the MRR (hereafter termed ‘divergent criteria’):

- The overall length of the reading passage was ranked by 53% of the participants, but with an MRR of 9.14. This criterion thus seems important to consider by the majority of participants, but not as a priority.

- Taking into account the vowel-to-vowel coarticulation was ranked by only 20% of the participants, but with an MRR of 6.5. Therefore, although it was not selected by the majority, it seemed to be quite an important criterion for a subgroup of the experts.

Furthermore, 7.5% (3/40) of the experts add in their comments that the text should also contain pseudo-words (which were not initially included in the available criteria). This element was added to round 2 for validation by the entire panel.

**[FIGURE 2 NEAR HERE]**

### ***Round 2***

Participants then rated a set of the most and best ranked criteria from the first round, as well as two additional statements regarding the ‘divergent’ criteria and one regarding pseudowords. The results of this round are presented in Table 3 as percent agreement for each of the statements grouped by target concepts (asterisks indicate that the consensus threshold was not reached). Consensus was reached for three of the four statements (>75% agreement among raters), with high agreement (>90%) for two of them. For the set of eight essential criteria, only one expert did not indicate agreement, stating that criteria 7 and 8 were not important to him because they can be tested in other ways.

**[TABLE 3 NEAR HERE]**

Only 50% of the experts agreed with the statement about pseudowords. Of these, 41% were strictly against the inclusion of pseudowords in a text (mainly because they are unnatural and could therefore have an impact on prosody), while another 41% specified that they were allowed but not necessary (depending on the purpose of the text).



### ***Round 3***

A third round was necessary to clarify this statement about the inclusion of pseudowords. Indeed, reading the comments revealed that there seemed to be some ambiguities in the wording. Therefore, the statement was rephrased and clarified in the third round: ‘The reading passage should contain pseudo-words in order to assess acoustic-phonetic decoding, provided that the integration is done in a way that best respects the natural character of the reading (e.g., pseudowords as proper names or incantations).’

In this final round, with 73% agreement (24/33), respondents still did not reach a consensus. Those who still disagree mainly stand by their opinion that pseudowords are not natural enough; one recommends using low-predictability sentences instead. Given this lack of consensus, it was concluded that pseudowords should not be systematically included in a text for speech and voice assessment. The focus (at least for perceptual assessment) should be on comprehensibility, in a more functional and ecological way. The assessment of intelligibility, on the other hand, hence requires a specific and distinct task, using unpredictable stimuli (pseudowords, minimal word pairs or unpredictable sentences).

### ***Final outcome***

This international Delphi consensus study allowed us to identify **ten main criteria** among the original 25 to take into account when creating a reading passage for the assessment of speech and voice, according to the expert panel. This set contains five phoneme-level, two word-level, two sentence-level criteria and one global-level criterion, as shown in

*Phoneme-level criteria:*

- I. Complete phonemic inventory (i.e. the text includes all of the target language's phonemes at least once);
- II. Phonetic balance (i.e. the included phonemes occur at approximately the same frequency at which they occur in ordinary conversation in the target language);
- III. Inclusion of (bi- and triconsonantal) clusters;
- IV. Taking into account the positions of phonemes in the words (beginning/middle/end), as well as their phonemic context;
- V. Taking into account vowel-to-vowel coarticulation (e.g. word pair /p a p a / - / p a p i /);

*Word-level criteria:*

- VI. Control of the articulatory/phonetic complexity of words;
- VII. Repeated occurrence of some words (error consistency, apraxia vs. dysarthria);

*Sentence-level criteria:*

- VIII. Presence of various intonation patterns (questions, statements, contrastive emphasis...);
- IX. Inclusion of various sentence lengths (breathing support, fatigability...);

*Global-level criteria:*

- X. Control of the overall length of the reading passage.

[FIGURE 3 NEAR HERE]

## **Discussion**

Reading passages are very commonly used, both in clinical practice and in research, as they provide predictable speech material with physiological demands resembling those of conversational speech (Mendes et al., 2012). The choice of the criteria to take into account when creating reading passages for speech and voice assessment in adolescents and adults may evidently vary depending on specific aims. However, the goal of this Delphi study was to provide a general baseline containing the criteria that the international expert panel judged to be essential for an overall assessment of speech and voice functions.

### ***Expert panel***

While a panel of 12 to 15 experts has been reported in numerous studies (McPherson et al., 2018), 40 experts participated in the first round of the present study. Furthermore, a low

dropout rate of 17% from the first to the third round was observed, while rates of 20 to 30% are usually expected (Chalmers & Armour, 2019). Though the sample size is thus satisfying, even more importantly, the characteristics of the expert panel are interesting, as it includes participants from various speech-related fields, backgrounds, and cultural and linguistic contexts. Furthermore, the expert panel profiles remained constant throughout the Delphi process (see Appendix A). Hence, the Delphi panel was considered as satisfying both in size and composition to address the main objective of this study.

### *Criteria for the creation of standard reading passages*

The resulting set validated by this panel contains five phoneme-level, two word-level, two sentence-level criteria and one global-level criterion. We will now discuss possible reasons behind this selection.

First, we hypothesize that the preponderance of phoneme-level criteria originates from the focus of the reading passage on speech rather than on language assessment. Speech intelligibility, for example, is indeed mostly assessed with a focus on segmental units, while prosody also plays a part (Pommée, Balaguer, Mauclair, et al., 2021b). Hence, prosody was consistently selected by the expert panel in addition to the five phoneme-level criteria. Lexical frequency, syntactical complexity and the passage's readability, on the other hand, are criteria that relate more to linguistic and cognitive rather than to speech production skills. While a speech-related rationale exists (see Appendix C), their relevance might thus be less straightforward, which could explain that they were not considered essential.

A second hypothesis regarding some of the selected criteria is that they are the most well-known and documented in the literature. The phonemic inventory and the phonetic balance, for example, are widely used in various speech production materials. Similarly, taking into account the positions of phonemes and their phonemic context is consistent with

the extensive literature regarding the effect of coarticulation on speech production and perception (e.g. [Katz et al., 1991; Liberman et al., 1967; Nguyen, 2001; Suomi, 1985; Van Son & Pols, 1999]). What is more surprising, however, is that the repetition of cardinal vowels was not considered essential, while vowel space measures are the most commonly investigated in speech research (Pommée, Balaguer, Piquier, et al., 2021). Vowel-to-vowel coarticulation, on the other hand, although it is not yet widely documented and studied, was considered as very important by a subset of experts. No comment nor common participant characteristic could be identified to explain the importance of this criterion for this subset of experts.

A third hypothesis refers to the properties of the panel of experts. The two criteria more specifically relating to fluency disorders (ratio of content/function words; initial voiceless stops, clusters and unvoiced-voiced phoneme sequences) were not considered as essential. This can be explained by the fact that only two of the experts were exclusively specialized in fluency disorders, with a total of 33% of the experts being at least partly active in this field and an average percentage of patients with fluency disorders of only about 5 %. Similarly, the inclusion of a 100% voiced sentence would be most interesting for experts specialized in the field of voice and phonation. While 58% of the experts state that they are at least partly active in the field of voice, none of them exclusively targets this field, and the average percentage of dysphonic patients (structural, functional and neurogenic voice disorders) only reaches about 15%. Hence, the non-selection of these criteria could be explained by the fact that the expert panel is mostly composed of professionals active in the field of speech disorders. The expert panel properties as well as their patient populations also explain the inclusion of criteria mostly relating to dysarthria and apraxia (e.g. repeated occurrence of some words for error consistency, articulatory complexity of words, inclusion of consonant clusters).

Overall, while using a consensus procedure allows to identify the main points on which a panel of experts agree, its use for the creation of new standard reading passages also has a pitfall regarding potentially under-represented subgroups of experts. Indeed, some criteria (e.g. vowel-to-vowel articulation, inclusion of a 100% voiced sentence and inclusion of pseudo-words), while discarded by most experts, nonetheless seem to be of interest for specific purposes. These niche criteria should therefore not be dismissed at first hand, but rather be considered as secondary as compared to the main set. If they can be included without impeding the main criteria or the readability, a new reading passage constructed with both the main and the niche criteria will benefit the largest number of potential users.

### ***Construction of a new reading passage***

How to proceed now that the minimal set of criteria for the creation of standard reading passages has been identified? This set is meant to be used as a general guideline. We applied these guidelines to create a new standard reading passage in French for patients aged at least 12 years (taking into account reading proficiency and allowing for a non-infantilizing reading content for adults). The description of the construction of this text goes beyond the scope of this article. However, to offer an idea of the process, its steps will now briefly be described to provide a baseline, e.g. for application in other languages.

The first step for the creation of our French reading passage was to set up an international and *multidisciplinary working group* of experts from Belgium, France and Quebec. This working group gathered to discuss existing reading passages and why these did not satisfy researchers' and clinicians' needs.

To get a broader view of these needs in the French-speaking community with regard to the speech and voice assessment, we carried out an *online survey*. This survey allowed us to

identify the measures to be applied on the future text and the target populations, and to define the specific objective and scope of the desired outcome.

Once the aim was set, we carried out a *review of existing reading passages* in French. This step allows identifying the most suitable existing reading passage, if applicable. Considering the minimal set of criteria described above, it is important at this stage to decide whether to use an existing passage or to start from scratch. If reading passages exist for which the construction criteria have been described, these might provide a good working basis. Otherwise, starting from scratch might prove easier and timesaving.

In our case, the recent text from the MonPaGe protocol (Pernon et al., 2020) proved to be resourceful and rather well described. We thus *contacted the authors* to obtain permission to use it as a working base. This text was then adapted to meet the minimal set of criteria, as well as *additional criteria* based on the needs identified by our working group and taking into account the *linguistic and cultural specificities* of the French-speaking context.

Furthermore, we decided to control as many criteria as possible in a first part of the text to allow for a *short reading*. This way, if a reading of the whole text is not possible because of time constraints or the patient's fatigability, this short reading will nevertheless allow extracting some parameters to analyse the patient's speech.

### **Limits and future perspectives**

The minimal set of criteria provided in this study applies to the creation of standardized reading passages for the assessment of speech and voice in patients with sufficient reading proficiency. Reading passages are not ideal for the assessment of speech and voice in young children (e.g. for the early identification of developmental speech sound disorders) or in illiterate adults, as reading difficulties might affect the speech production.

The expert panel that participated in this study slightly weights toward neurogenic adult speech disorders (dysarthria and apraxia). Criteria that are more specific to the

assessment of fluency and voice disorders might therefore have been discounted and could be considered in addition to the main criteria set.

It is also to be noted that the minimal set of criteria provides a guideline that mainly applies to Indo-European Romance and Germanic languages such as English, French and German. Other language groups, such as Indo-European Balto-Slavic or Sino-Tibetan languages (including Chinese), are not or only little represented in the expert panel. Separate studies may be necessary to identify the most important criteria that apply to the specificities of these languages. Furthermore, this set is not adapted for specific research aims, for which particular criteria might be needed (see for example (Kuo & Weismer, 2016; van Zundert et al., 1998)).

## **Conclusion**

Despite their importance for the assessment of speech and voice, available standard reading passages present various limitations. To allow for the creation in various languages of new texts for standard and ecological overall speech and voice assessment, an international Delphi consensus survey was carried out to identify the most important criteria to take into account. This study results in a minimal set of ten criteria which were selected by a majority of the experts and were rated as most important. It contains five phoneme-level, two word-level, two sentence-level criteria and one global-level criterion. It can be used as a general guideline for the creation of standard reading passages in Indo-European Romance and Germanic languages such as English, French and German.

## **Acknowledgments**

The authors thank all the volunteers who gave their time to participate in this study, as well as Prof. Renée Speyer (University of Oslo) for her guidance regarding the Delphi methodology.

This project was supported by the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie Grant 766287.

### **Declaration of interest**

The authors declare that there is no conflict of interest.

### **References**

- Adams, M. R., & Reis, R. (1974). Influence of the onset of phonation on the frequency of stuttering: A replication and reevaluation. *Journal of Speech and Hearing Research*, 17(4), 752–754. <https://doi.org/10.1044/jshr.1704.752>
- Aichert, I., & Ziegler, W. (2004). Syllable frequency and syllable structure in apraxia of speech. *Brain and Language*, 88(1), 148–159. [https://doi.org/10.1016/S0093-934X\(03\)00296-7](https://doi.org/10.1016/S0093-934X(03)00296-7)
- Allison, K. M., & Hustad, K. C. (2014). Impact of sentence length and phonetic complexity on intelligibility of 5-year-old children with cerebral palsy. *International Journal of Speech-Language Pathology*, 16(4), 396–407. <https://doi.org/10.3109/17549507.2013.876667>
- Allison, K. M., Yunusova, Y., & Green, J. R. (2019). Shorter sentence length maximizes intelligibility and speech motor performance in persons with dysarthria due to Amyotrophic Lateral Sclerosis. *American Journal of Speech-Language Pathology*, 28(1), 96–107. [https://doi.org/10.1044/2018\\_AJSLP-18-0049](https://doi.org/10.1044/2018_AJSLP-18-0049)
- Auzou, P., & Rolland-Monnoury, V. (2006). *BECD 2006 - Batterie d'Évaluation Clinique de la Dysarthrie*. Orthoédition.
- Auzou, P., & Rolland-Monnoury, V. (2019). *Batterie d'Évaluation Clinique de la Dysarthrie - BECD 2019*. Orthoédition.



- Barreto, S. dos S., & Ortiz, K. Z. (2020). Speech intelligibility in dysarthrias: Influence of utterance length. *Folia Phoniatrica et Logopaedica*, 72(3), 202–210.  
<https://doi.org/10.1159/000497178>
- Beverly, D., Cannito, M. P., Chorna, L., Wolf, T., Suiter, D. M., & Bene, E. R. (2010). Influence of stimulus sentence characteristics on speech intelligibility scores in hypokinetic dysarthria. *Journal of Medical Speech - Language Pathology*, 18(4).
- Birko, S., Dove, E. S., Özdemir, V., & Dalal, K. (2015). Evaluation of nine consensus indices in delphi foresight research and their dependency on delphi survey characteristics: A simulation study and debate on delphi design and interpretation. *PLoS ONE*, 10(8), 1–14.  
<https://doi.org/10.1371/journal.pone.0135162>
- Boering, G., Hoeksema, P. E., Huffstadt, A. J. C., Moolenaar-Bijl, A. J., Ritsema Van Eck, C. R., & Visser, H. K. A. (1961). *Over de behandeling van patienten met cheilognathopalatoschisis*. Academisch Ziekenhuis.
- Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, 16(6), 451–459. <https://doi.org/10.1111/j.0956-7976.2005.01556.x>
- Boucher, V. J., & Lalonde, B. (2015). Effects of the growth of breath capacities on mean length of utterances: How maturing production processes influence indices of language development. *Journal of Phonetics*, 52, 58–69.  
<https://doi.org/10.1016/j.wocn.2015.04.005>
- Brown, S. F. (1945). The loci of stutterings in the speech sequence. *Journal of Speech Disorders*, 10(3), 181–192. <https://doi.org/10.1044/jshd.1003.181>
- Camargo, Z., Madureira, S., Reis, N., Rilliard, A., Camargo, Z., Madureira, S., Reis, N., Rilliard, A., Camargo, Z., Madureira, S., Reis, N., & Rilliard, A. (2019). The phonetic

approach of voice qualities : challenges in corresponding perceptual to acoustic descriptions. In *Subsidia. Tools and resources for speech sciences*.

- Chalmers, J., & Armour, M. (2019). The delphi technique. *Handbook of Research Methods in Health Social Sciences*, 715–735. [https://doi.org/10.1007/978-981-10-5251-4\\_99](https://doi.org/10.1007/978-981-10-5251-4_99)
- Chen, H.-C., Vaid, J., Boas, D. A., & Bortfeld, H. (2011). Examining the phonological neighborhood density effect using near infrared spectroscopy. *Human Brain Mapping*, 32(9), 1363–1370. <https://doi.org/10.1002/hbm.21115>
- Cole, R., Yan, Y., Mak, B., & Fanty, M. (1996). The contribution of consonants versus vowels to word recognition in fluent speech. *IEE Proceedings - Vision, Image, and Signal Processing*, 2, 853–856. <https://doi.org/10.1121/1.417028>
- Cunningham, B. J., Kwok, E.-, Turkstra, L., & Oram Cardy, J. (2019). Establishing consensus among community clinicians on how to categorize and define preschoolers' speech and language impairments at assessment. *Journal of Communication Disorders*, 82(November-December), 105925. <https://doi.org/10.1016/j.jcomdis.2019.105925>
- D'Alessandro, D., & Fougeron, C. (2018). Réduction de la coarticulation et vieillissement. *Journées d'Étude Sur La Parole*, 410–418. <https://hal.archives-ouvertes.fr/hal-02436286>
- D'Alessandro, D., Pernon, M., Fougeron, C., Laganaro, M., Alessandro, D., Pernon, M., Fougeron, C., Laganaro, M., & Vtov, A. (2019). Anticipatory VtoV coarticulation in French in several motor speech disorders. *Phonetics and Phonology in Europe*. <https://hal.archives-ouvertes.fr/hal-02427864>
- Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research*, 12(2), 246–269. <https://doi.org/10.1044/jshr.1202.246>
- Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders* (3rd ed.). W.B. Saunders Company.

- Darley, Frederic L, Aronson, A. E., & Brown, J. R. (1969). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research*, *12*(3), 462–496. <https://doi.org/10.1044/jshr.1203.462>
- Daudet, A. (1869). La chèvre de monsieur Seguin. In J. Hetzel (Ed.), *Lettres de mon moulin* (pp. 37–50).
- de Maupassant, G. (1882). Pierrot. *Le Gaulois*.
- Denman, D., Kim, J. H., Munro, N., Speyer, R., & Cordier, R. (2019). Describing language assessments for school-aged children: A Delphi study. *International Journal of Speech-Language Pathology*, *21*(6), 602–612. <https://doi.org/10.1080/17549507.2018.1552716>
- Di Cristo, A. Di. (2012). Le pouvoir de la prosodie ou la revanche de Cendrillon. In L. Baqué & M. Estrada (Eds.), *L'homme communicant* (pp. 95–114). CIPA.
- Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401–409. <https://doi.org/10.1016/j.jclinepi.2013.12.002>
- Duffy, J. R. (2013). *Motor Speech Disorders: Substrates, differential diagnosis, and management* (3rd ed.). Elsevier.
- Dufour, S., & Frauenfelder, U. H. (2010). Phonological neighbourhood effects in French spoken-word recognition. *Quarterly Journal of Experimental Psychology*, *63*(2), 226–238. <https://doi.org/10.1080/17470210903308336>
- Dworzynski, K., Howell, P., Au-Yeung, J., & Rommel, D. (2004). Stuttering on function and content words across age groups of German speakers who stutter. *Journal of Multilingual Communication Disorders*, *2*(2), 81–101. <https://doi.org/10.1080/14769670310001625354>
- Fairbanks, G. (1960). *Voice and articulation drillbook* (2nd ed.). Harper & Row.

- Ferguson, S. H., & Quené, H. (2014). Acoustic correlates of vowel intelligibility in clear and conversational speech for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *135*(6), 3570–3584.  
<https://doi.org/10.1121/1.4874596>
- Fletcher, S. G. (1972). Contingencies for bioelectronic modification of nasality. *The Journal of Speech and Hearing Disorders*, *37*(3), 329–346. <https://doi.org/10.1044/jshd.3703.329>
- Fougeron, C., Delvaux, V., Ménard, L., & Laganaro, M. (2019). The MonPaGe\_HA database for the documentation of spoken French throughout adulthood. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 4301–4306.
- Fredouille, C., Ghio, A., Laaridh, I., Lalain, M., Woisard, V., Fredouille, C., Ghio, A., Laaridh, I., Lalain, M., & Woisard, V. (2019). Acoustic-phonetic decoding for speech intelligibility evaluation in the context of head and neck cancers. *International Congress of Phonetic Sciences*, 3051–3055.
- Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language*, *66*(4), 789–806. <https://doi.org/10.1016/j.jml.2011.11.006>
- Galluzzi, C., Bureca, I., Guariglia, C., & Romani, C. (2015). Phonological simplifications, apraxia of speech and the interaction between phonological and phonetic processing. *Neuropsychologia*, *71*, 64–83. <https://doi.org/10.1016/j.neuropsychologia.2015.03.007>
- Gendrot, C., & Adda-Decker, M. (2005). Impact of duration on F1/F2 formant values of oral vowels: An automatic analysis of large broadcast news corpora in French and German. *9th European Conference on Speech Communication and Technology*, *1*, 2453–2456.
- Gurevich, N., & Scamihorn, S. L. (2017). Speech-language pathologists' use of intelligibility measures in adults with dysarthria. *American Journal of Speech-Language Pathology*, *26*(3), 873–892. [https://doi.org/10.1044/2017\\_AJSLP-16-0112](https://doi.org/10.1044/2017_AJSLP-16-0112)

- Haley, K. L., Cunningham, K. T., Eaton, C. T., & Jacks, A. (2018). Error consistency in acquired apraxia of speech with aphasia: Effects of the analysis unit. *Journal of Speech, Language, and Hearing Research, 61*(2), 210–226.  
[https://doi.org/10.1044/2017\\_JSLHR-S-16-0381](https://doi.org/10.1044/2017_JSLHR-S-16-0381)
- Hardcastle, B., & Tjaden, K. (2009). Coarticulation and speech impairment. *The Handbook of Clinical Linguistics, 506–524*. <https://doi.org/10.1002/9781444301007.ch32>
- Harmegnies, B. (1988). *Contribution à la caractérisation de la qualité vocale. Analyses plurielles de spectres moyens à long terme de parole*. Université de Mons-Hainaut.
- Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *The Journal of the Acoustical Society of America, 109*(2), 748–763. <https://doi.org/10.1121/1.1337959>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288.  
<https://doi.org/10.1177/1049732305276687>
- Huinck, W. J., Van Lieshout, P. H. H. M., Peters, H. F. M., & Hulstijn, W. (2004). Gestural overlap in consonant clusters: Effects on the fluent speech of stuttering and non-stuttering subjects. *Journal of Fluency Disorders, 29*(1), 3–25.  
<https://doi.org/10.1016/j.jfludis.2003.09.001>
- International Phonetic Association. (1999). *Handbook of the International Phonetic Association*. Cambridge University Press.
- Jakielski, K. J. (1998). *Motor organization in the acquisition of consonant clusters* [University of Texas at Austin].  
<https://www.proquest.com/openview/e3c556421c6b8f8dbeebf23f42fe1fdd/1?pq-origsite=gscholar&cbl=18750&diss=y>
- Jakielski, K. J. (2000). Quantifying phonetic complexity in words: An experimental index.

*Annual Child Phonology Conference.*

- Jesus, L. M. T., Valente, A. R. S., & Hall, A. (2015). Is the Portuguese version of the passage “The North Wind and the Sun” phonetically balanced? *Journal of the International Phonetic Association*, 45(1), 1–11. <https://doi.org/10.1017/S0025100314000255>
- Katz, W. F. (2000). Anticipatory coarticulation and aphasia: Implications for phonetic theories. *Journal of Phonetics*, 28(3), 313–334. <https://doi.org/10.1006/jpho.2000.0118>
- Katz, W. F., Kripke, C., & Tallal, P. (1991). Anticipatory coarticulation in the speech of adults and young children: Acoustic, perceptual, and video data. *Journal of Speech, Language, and Hearing Research*, 34(6), 1222–1232. <https://doi.org/10.1044/jshr.3406.1222>
- Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology*, 18(2), 124–132. [https://doi.org/10.1044/1058-0360\(2008/08-0017\)](https://doi.org/10.1044/1058-0360(2008/08-0017))
- Kent, R. D., Kent, J. F., Weismer, G., Martin, R. E., Sufit, R. L., Brooks, B. R., & Rosenbek, J. C. (1989). Relationships between speech intelligibility and the slope of second-formant transitions in dysarthric subjects. *Clinical Linguistics & Phonetics*, 3(4), 347–358. <https://doi.org/10.3109/02699208908985295>
- Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 122(4), 2365–2375. <https://doi.org/10.1121/1.2773986>
- Kim, H., Martin, K., Hasegawa-Johnson, M., & Perlman, A. (2010). Frequency of consonant articulation errors in dysarthric speech. *Clinical Linguistics and Phonetics*, 24(10), 759–770. <https://doi.org/10.3109/02699206.2010.497238>

- Kuo, C., & Weismer, G. (2016). Vowel reduction across tasks for male speakers of American English. *The Journal of the Acoustical Society of America*, *140*(1), 369–383.  
<https://doi.org/10.1121/1.4955310>
- Kuruvilla-Dugdale, M., Custer, C., Heidrick, L., Barohn, R., & Govindarajan, R. (2018). A phonetic complexity-based approach for intelligibility and articulatory precision testing: A preliminary study on talkers with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research : JSLHR*, *61*(9), 2205–2214.  
[https://doi.org/10.1044/2018\\_JSLHR-S-17-0462](https://doi.org/10.1044/2018_JSLHR-S-17-0462)
- Kuruvilla-Dugdale, M., Dietrich, M., McKinley, J. D., & Deroche, C. (2020). An exploratory model of speech intelligibility for healthy aging based on phonatory and articulatory measures. *Journal of Communication Disorders*, *87*(March), 105995.  
<https://doi.org/10.1016/j.jcomdis.2020.105995>
- Laganaro, M., Fougeron, C., Pernon, M., Levêque, N., Borel, S., Fournet, M., Catalano Chiuvé, S., Lopez, U., Trouville, R., Ménard, L., Burkhard, P. R., Assal, F., & Delvaux, V. (2021). Sensitivity and specificity of an acoustic- and perceptual-based tool for assessing motor speech disorders in French: the MonPaGe-screening protocol. *Clinical Linguistics & Phonetics*, 1–16. <https://doi.org/10.1080/02699206.2020.1865460>
- Lee, H., Gambette, P., Barkat-defradas, M., Lee, H., Gambette, P., & Barkat-defradas, M. (2014). iPhocomp : calcul automatique de l'indice de complexité phonétique de Jakielski. *Journées d'Étude Sur La Parole*, 622–630.
- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.  
<https://doi.org/10.1037/h0020279>
- Linstone, H. A., & Turoff, M. (2002). *The Delphi method: Techniques and applications* (M. Turoff & H. A. Linstone (eds.)). New Jersey Institute of Technology.

- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*(1), 1–36. <https://doi.org/10.1097/00003446-199802000-00001>
- Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [j]-[s] distinction. *Perception & Psychophysics, 28*(3), 213–228. <https://doi.org/10.3758/BF03204377>
- Martel-Sauvageau, V., & Tjaden, K. (2017). Vocalic transitions as markers of speech acoustic changes with STN-DBS in Parkinson's disease. *Journal of Communication Disorders, 70*(12), 1–11. <https://doi.org/10.1016/j.jcomdis.2017.10.001>
- Martens, H., & Nuffelen, G. Van. (2010). De ontwikkeling van een fonetischgebalanceerde standaardtekst. *Logopedie, 31*–36.
- Maurová Paillereau, N. (2016). Do isolated vowels represent vowel targets in French? An acoustic study on coarticulation. *SHS Web of Conferences, 27*, 09003. <https://doi.org/10.1051/shsconf/20162709003>
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86. [https://doi.org/10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0)
- McPherson, S., Reese, C., & Wendler, M. C. (2018). Methodology update: Delphi studies. *Nursing Research, 67*(5), 404–410. <https://doi.org/10.1097/NNR.0000000000000297>
- Mendes, A. ., Da Costa, A., Martins, A. ., Fernandes, A. F. ., Do Rosario Vicente, S. ., & Freitas, T. . S. (2012). Contributions to the elaboration of a phonetically balanced text for the European-Portuguese. *Revista CEFAC, 14*(5), 910–917.
- Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research, 47*(5), 1048–1058. [https://doi.org/10.1044/1092-4388\(2004/078\)](https://doi.org/10.1044/1092-4388(2004/078))
- Nguyen, N. (2001). Rôle de la coarticulation dans la reconnaissance des mots. *Annee*



*Psychologique*, 101(1), 125. <https://doi.org/10.3406/psy.2001.29719>

- Owren, M. J., & Cardillo, G. C. (2006). The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *The Journal of the Acoustical Society of America*, 119(3), 1727–1739. <https://doi.org/10.1121/1.2161431>
- Patel, R., Connaghan, K., Franco, D., Edsall, E., Forgit, D., Olsen, L., Ramage, L., Tyler, E., & Russell, S. (2013). “The Caterpillar”: A novel reading passage for assessment of motor speech disorders. *American Journal of Speech-Language Pathology*, 22(1), 1–9. [https://doi.org/10.1044/1058-0360\(2012/11-0134\)](https://doi.org/10.1044/1058-0360(2012/11-0134))
- Pernon, M., Lévêque, N., Delvaux, V., Assal, F., Borel, S., Fougeron, C., Trouville, R., & Laganaro, M. (2020). MonPaGe, un outil de screening francophone informatisé d'évaluation perceptive et acoustique des troubles moteurs de la parole (dysarthries, apraxie de la parole). *Rééducation Orthophonique*, 281, 169–198.
- Pommée, T., Balaguer, M., Mauclair, J., Pinquier, J., & Woisard, V. (2021a). Assessment of adult speech disorders: current situation and needs in French-speaking clinical practice. *Logopedics Phoniatrics Vocology*, 0(0), 1–15. <https://doi.org/10.1080/14015439.2020.1870245>
- Pommée, T., Balaguer, M., Mauclair, J., Pinquier, J., & Woisard, V. (2021b). Intelligibility and comprehensibility: A Delphi consensus study. *International Journal of Language & Communication Disorders*. <https://doi.org/10.1111/1460-6984.12672>
- Pommée, T., Balaguer, M., Pinquier, J., Mauclair, J., Woisard, V., & Speyer, R. (2021). Relationship between phoneme-level spectral acoustics and speech intelligibility in healthy speech: a systematic review. *Speech, Language and Hearing*, 24(2), 105–132. <https://doi.org/10.1080/2050571X.2021.1913300>
- Pouchoulin, G., Fredouille, C., Ghio, A., Azzarello, M., Giovanni, A., Pouchoulin, G., Fredouille, C., Ghio, A., & Azzarello, M. (2007). Modélisation statistique et

informations pertinentes pour la caractérisation des voix pathologiques (dysphonies).

*Journées d'Étude Sur La Parole*. [https://dumas.ccsd.cnrs.fr/SANTE\\_PUB\\_INSERM/hal-00136742v1](https://dumas.ccsd.cnrs.fr/SANTE_PUB_INSERM/hal-00136742v1)

Powell, T. (2006). A comparison of English reading passages for elicitation of speech samples from clinical populations. *Clinical Linguistics and Phonetics*, 20(2–3), 91–97.

<https://doi.org/10.1080/02699200400026488>

Reilly, J., & Fisher, J. L. (2012). Sherlock Holmes and the strange case of the missing attribution: A historical note on “The Grandfather Passage.” *Journal of Speech, Language, and Hearing Research*, 55(1), 84–88. [https://doi.org/10.1044/1092-4388\(2011/11-0158\)](https://doi.org/10.1044/1092-4388(2011/11-0158))

Révis, J. (2004). *L'analyse perceptive des dysphonies : approche phonétique de l'évaluation vocale* [Université Aix-Marseille II]. <http://www.theses.fr/2004AIX20675>

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010).

Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research*, 53(2), 333–349. [https://doi.org/10.1044/1092-4388\(2009/08-0183\)](https://doi.org/10.1044/1092-4388(2009/08-0183))

Rosenbek, J. C., Kent, R. D., & Lapointe, L. L. (1984). Apraxia of speech: An overview and some perspectives. In J. C. Rosenbek, M. R. McNeil, & A. E. Aronson (Eds.), *Apraxia of speech: Physiology, acoustics, linguistics, management* (pp. 1–72). College-Hill Press.

Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant centralization ratio: A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research*, 53(1), 114–125. [https://doi.org/10.1044/1092-4388\(2009/08-0184\)](https://doi.org/10.1044/1092-4388(2009/08-0184))

Sasisekaran, J. (2014). Exploring the link between stuttering and phonology: A review and implications for treatment. *Seminars in Speech and Language*, 35(2), 95–113.

<https://doi.org/10.1055/s-0034-1371754>

Skodda, S., Grönheit, W., & Schlegel, U. (2012). Impairment of vowel articulation as a possible marker of disease progression in Parkinson's disease. *PLoS ONE*, 7(2).

<https://doi.org/10.1371/journal.pone.0032132>

Staiger, A., Finger-Berg, W., Aichert, I., & Ziegler, W. (2012). Error variability in apraxia of speech: A matter of controversy. *Journal of Speech, Language, and Hearing Research*, 55(5). [https://doi.org/10.1044/1092-4388\(2012/11-0319\)](https://doi.org/10.1044/1092-4388(2012/11-0319))

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: an acoustical study. *Journal of Speech and Hearing Research*, 6(2), 111–128.

<https://doi.org/10.1044/jshr.0602.111>

Suomi, K. (1985). The vowel-dependence of gross spectral cues to place of articulation of stop consonants in CV syllables. *Journal of Phonetics*, 13(3), 267–285.

[https://doi.org/10.1016/s0095-4470\(19\)30759-4](https://doi.org/10.1016/s0095-4470(19)30759-4)

Teston, B. (2004). L'évaluation instrumentale des dysphonies : Etat actuel et perspectives. In A. Giovanni (Ed.), *Le bilan d'une dysphonie* (pp. 105–169). Solal.

Van de Weijer, J., & Slis, I. (1991). Nasaliteitsmeting met de Nasometer. *Logopedie En Foniatrie*, 63, 97–101.

Van Riper, C. (1963). *Speech correction: Principles and methods* (4th ed.). Prentice-Hall.

Van Son, R. J. J. H., & Pols, L. C. W. (1999). Acoustic description of consonant reduction. *Speech Communication*, 28(2), 125–140. [https://doi.org/10.1016/S0167-6393\(99\)00009-](https://doi.org/10.1016/S0167-6393(99)00009-6)

6

van Zundert, M., Rietveld, T., & Mugge, A. M. (1998). Normgegevens voor het gebruik van de nasometer bij jonge kinderen. *Stem-, Spraak- En Taalpathologie*, 7(4), 237–247.

Vasilescu, I., Candea, M., & Adda-decker, M. (2004). Hésitations autonomes dans 8 langues : une étude acoustique et perceptive. *Colloque MIDL 2004, Modélisations Pour*

*l'identification Des Langues et Des Variétés Dialectales, 29-30 Novembre, 29–30.*

- Vitevitch, M. S. (2002). Influence of onset density on spoken-word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 270–278.  
<https://doi.org/10.1037/0096-1523.28.2.270>
- von der Gracht, H. A. (2012). Consensus measurement in Delphi studies. Review and implications for future quality assurance. *Technological Forecasting and Social Change, 79*(8), 1525–1536. <https://doi.org/10.1016/j.techfore.2012.04.013>
- Weismer, G., Laures, J. S., Jeng, J., Kent, R. D., & Kent, J. F. (2000). Effect of speaking rate manipulations on acoustic and perceptual aspects of the dysarthria in amyotrophic lateral sclerosis. *Folia Phoniatrica et Logopaedica, 53*(706), 201–219.
- Wells, G. B. (1983). A feature analysis of stuttered phonemes. *Journal of Fluency Disorders, 8*(2), 119–124. [https://doi.org/10.1016/0094-730X\(83\)90024-4](https://doi.org/10.1016/0094-730X(83)90024-4)
- Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[S] boundary. *Journal of the Acoustical Society of America, 69*(1), 275–282.  
<https://doi.org/10.1121/1.385348>
- Wingate, M. E., & Howell, P. (2002). Foundations of stuttering. *The Journal of the Acoustical Society of America, 112*(4), 1229–1231. <https://doi.org/10.1121/1.1501910>
- Woisard, V., Espesser, R., Ghio, A., & Duez, D. (2013). De l' intelligibilité à la compréhension de la parole, quelles mesures en pratique clinique ? *Rev Laryngol Otol Rhinol, 134*(1), 27–33.
- Wolk, L., Blomgren, M., & Smith, A. B. (2000). The frequency of simultaneous disfluency and phonological errors in children: A preliminary investigation. *Journal of Fluency Disorders, 25*(4), 269–281. [https://doi.org/10.1016/S0094-730X\(00\)00076-0](https://doi.org/10.1016/S0094-730X(00)00076-0)
- Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition,*

31(6), 1385–1397. <https://doi.org/10.1037/0278-7393.31.6.1385>

Yorkston, K. M., Beukelman, D. R., & Tice, R. (1996). *Sentence intelligibility test*. Tice Technologies.

Ziegler, W., & Aichert, I. (2015). How much is a word? Predicting ease of articulation planning from apraxic speech error patterns. *Cortex*, 69, 24–39.

<https://doi.org/10.1016/j.cortex.2015.04.001>

## Appendices

### Appendix A – Description of the expert panel

Round	1	2	3
Participants	N= 40	N= 34	N= 33
Fields of activity	N (% <sup>1</sup> )	N (%)	N (%)
Speech disorders (SD)	38 (95%)	32 (94.1%)	31 (93.9%)
SD only	13 (32.5%)	11 (32.4%)	11 (33.3%)
SD & voice/fluency	25 (62.5%)	21 (61.8%)	20 (60.6%)
Fluency disorders only	2 (5%)	2 (5.9%)	2 (6.1%)
Profession <sup>2</sup>			
Speech and language pathologist	29 (72.5%)	24 (70.6%)	23 (69.7%)
Linguist	9 (22.5%)	8 (23.5%)	8 (24.2%)
ENT/phoniatrician	8 (20%)	7 (20.6%)	7 (21.2%)
Computer scientist	7 (17.5%)	7 (20.6%)	7 (21.2%)
Psychologist/neuropsychologist	2 (5%)	2 (5.9%)	2 (6.1%)
Audiologist	2 (5%)	2 (5.9%)	2 (6.1%)
Neuroscientist	1 (2.5%)	1 (2.9%)	1 (3%)
Other (TEFL, performance/acting/singing)	2 (5%)	2 (5.9%)	2 (6.1%)
Years of experience in speech/voice			
>20 years	13 (32.5%)	12 (35.3%)	12 (36.4%)
15-20 years	6 (15%)	5 (14.7%)	5 (15.2%)
10-15 years	4 (10%)	3 (8.8%)	3 (9.1%)

5-10 years	10 (25%)	8 (23.5%)	8 (24.2%)
< 5 years	7 (17.5%)	6 (17.6%)	5 (15.2%)
<b>Professional activity</b>			
<b>First rank (main activity)</b>			
Research	14 (35%)	14 (41.2%)	14 (42.4%)
Clinical	13 (32.5%)	10 (29.4%)	9 (27.3%)
Academic	11 (27.5%)	9 (26.5%)	9 (27.3%)
Industrial	2 (5%)	1 (2.9%)	1 (3%)
Combinations	34 (85%)	32 (94.1%)	31 (93.9%)
Clinical activity only	4 (10%)	3 (8.8%)	2 (6.1%)
Clinical activity & Research	21 (53%)	18 (52.9%)	18 (54.5%)
Research & Academics	28 (70%)	26 (76.5%)	26 (78.8%)
<b>Level of education</b>			
Third cycle (PhD)	23 (57.5%)	21 (61.8%)	21 (63.6%)
Second cycle (Masters)	16 (40%)	13 (38.2%)	12 (36.4%)
of which PhD students	3 (7.5%)	3 (8.8%)	3 (9.1%)
First cycle (BA, BSc...)	1 (2.5%)	0 (0%)	0 (0%)
<b>Year of diplamation</b>			
Mean (std)	2009.6 (7.88)	2010.3 (7.72)	2010.1 (7.72)
Median (IQR)	2012 (11.5)	2014 (12)	2014 (12)
<b>Academic title (US system equivalentents)</b>			
Associate Professor	16 (40%)	15 (44.1%)	15 (45.5%)
No academic activity	8 (20%)	6 (17.6%)	5 (15.2%)

Lecturer	5 (12.5%)	4 (11.8%)	4 (12.1%)
No academic title (e.g. PhD)	3 (7.5%)	3 (8.8%)	3 (9.1%)
Professor	3 (7.5%)	3 (8.8%)	3 (9.1%)
Clinical Instructor	2 (5%)	0 (0%)	0 (0%)
Assistant Professor	1 (2.5%)	1 (2.9%)	1 (3%)
Professor Emeritus	1 (2.5%)	1 (2.9%)	1 (3%)
Senior Lecturer	1 (2.5%)	1 (2.9%)	1 (3%)
<hr/>			
Country			
France	7 (17.5%)	7 (20.6%)	7 (21.2%)
United Kingdom	7 (17.5%)	6 (17.6%)	6 (18.2%)
Germany	6 (15%)	5 (14.7%)	5 (15.2%)
Canada	3 (7.5%)	2 (5.9%)	2 (6.1%)
Australia	2 (5%)	2 (5.9%)	2 (6.1%)
Belgium	2 (5%)	2 (5.9%)	2 (6.1%)
Finland	2 (5%)	1 (2.9%)	0 (0%)
Malta	2 (5%)	2 (5.9%)	2 (6.1%)
Colombia	1 (2.5%)	1 (2.9%)	1 (3%)
Hong Kong	1 (2.5%)	0 (0%)	0 (0%)
India	1 (2.5%)	1 (2.9%)	1 (3%)
Netherlands	1 (2.5%)	1 (2.9%)	1 (3%)
Pakistan	1 (2.5%)	1 (2.9%)	1 (3%)
Spain	1 (2.5%)	1 (2.9%)	1 (3%)
Sweden	1 (2.5%)	1 (2.9%)	1 (3%)
Switzerland	1 (2.5%)	0 (0%)	0 (0%)
USA	1 (2.5%)	1 (2.9%)	1 (3%)



Main language <sup>2</sup>			
English	25 (62.5%)	22 (64.7%)	22 (66.7%)
French	12 (30%)	11 (32.4%)	11 (33.3%)
German	5 (12.5%)	4 (11.8%)	4 (12.1%)
Finnish	2 (5%)	1 (2.9%)	0 (0%)
Spanish	2 (5%)	2 (5.9%)	2 (6.1%)
Mandarin	1 (2.9%)	1 (2.9%)	1 (3%)
Cantonese	1 (2.9%)	0 (0%)	0 (0%)
Catala	1 (2.5%)	1 (2.9%)	1 (3%)
Dutch	1 (2.5%)	1 (2.9%)	1 (3%)
Hindi	1 (2.5%)	1 (2.9%)	1 (3%)
Kannada	1 (2.5%)	1 (2.9%)	1 (3%)
Malayalam	1 (2.5%)	1 (2.9%)	1 (3%)
Maltese	1 (2.5%)	1 (2.9%)	1 (3%)
Portuguese	1 (2.5%)	1 (2.9%)	1 (3%)
Swedish	1 (2.5%)	1 (2.9%)	1 (3%)
Urdu	1 (2.5%)	1 (2.9%)	1 (3%)
Main patient/study population <sup>3</sup>			
Age groups	Average % across participants		
Children (<10 years old)	24.8%	22.2%	20.6%
Adolescents (10-18 years old)	10.1%	7.8%	7.9%
Young adults (19-35 years old)	13.1%	11.5%	11.7%
Middle-aged adults (36-65 years old)	20.2%	22.3%	22.7%
Elderly (65+)	31.8%	36.2%	37.1%
Pathologies			

Neurodegenerative disorders	22.1%	25.7%	26.4%
Acquired neurological disorders	15.4%	17.1%	17.2%
Others (speech-/voice-related)	13.6%	8.8%	7.7%
Oncology	8.9%	10.4%	10.8%
Functional voice disorders	5.6%	5.4%	5.2%
Other structural speech deficits	5.2%	5.7%	5.9%
Fluency disorders	5.1%	4.7%	4.7%
Others (non speech-/voice-related)	4.9%	2.6%	2.1%
Structural voice disorders	4.6%	4.1%	4.1%
Hearing-impairment	4.5%	4.6%	4.7%
Neurogenic voice disorders	4.1%	4.6%	4.7%
Polyhandicap	3.8%	3.8%	3.9%
Neurological tumours	1.8%	2.1%	2.1%
Iatrogenesis	0.4%	0.4%	0.5%

<sup>1</sup> Due to rounding of numbers to one decimal, minimal deviations from 100% may occur.

<sup>2</sup> Combinations were possible (e.g. in countries with more than one official language).

<sup>3</sup> Each participant distributed 100% over all given categories; results are average percentages across participants for each category.

## Tables

Table 1. Measures and ratings participants want to be able to use on a standard reading passage, according to the results from Round 1.

<i>Measures and ratings</i>	<i>Mentions (/40)</i>
Suprasegmental measures	22
Voice-quality measures	21
Percent-correct identification scores	17
of which phoneme-level measures	11
Articulation/phonetic accuracy measures	17
of which using acoustic measures	10
Intelligibility (no specification as to how/at what level of granularity)	9
Subjective overall speech-related ratings	7

Table 2. Eight most and best ranked criteria among the initial set of 25

<i>Criterion</i>	<i>Mean rank (MR)</i>	<i>Rating frequency (RF)</i>	<i>Average rank excluding non-ratings (MRR)</i>
Complete phonemic inventory	9.5	68%	1.6
Phonetic balance	11.6	63%	3
Inclusion of various intonation patterns	11.9	73%	6.6
Control of the articulatory/phonetic complexity of words	12.6	68%	6.1
Taking into account the positions of the phonemes in the words, as well as their vowel/consonant context	15.6	50%	5.3
Inclusion of consonant clusters	15.9	50%	5.8
Repeated presence of some words	16.4	50%	6.7
Inclusion of various sentence lengths	16.6	50%	7.3

Table 3. Statements presented in the second round of the Delphi survey and percentages of agreement.

Concept	Statement	Agreement
<i>Pseudo-words</i>	- ‘The text should contain pseudowords to assess speech intelligibility.’	50% (17/34)*
<i>Essential criteria set</i>	- ‘The main criteria to take into account when creating a reading passage for speech and voice assessment are: <ol style="list-style-type: none"> <li>1. Complete phonemic inventory (i.e. the text includes all of the target language’s phonemes at least once).</li> <li>2. Phonetic balance (i.e. the included phonemes occur at approximately the same frequency at which they occur in ordinary conversation in the target language).</li> <li>3. Inclusion of various intonation patterns (questions, statements, contrastive emphasis...).</li> <li>4. Control of the articulatory/phonetic complexity of words.</li> <li>5. Taking into account the positions of the phonemes in the words (beg/mid/end), as well as their vowel/consonant context.</li> <li>6. Inclusion of consonant clusters.</li> <li>7. Repeated presence of some words (error consistency, apraxia vs dysarthria).</li> <li>8. Inclusion of various sentence lengths (breathing support, fatigability...).</li> </ol>	97% (33/34)
<i>Mixed opinion criteria</i>	- ‘The overall length of the reading passage has to be taken into account when creating a reading passage for speech and voice assessment.’	97% (33/34)

- 'Vowel-to-vowel coarticulation has to be taken into account 79% (27/34)  
when creating a reading passage for speech and voice  
assessment.'

---

\*consensus threshold not reached

## Figures

Figure 1. Flowchart of the Delphi process used in this study.

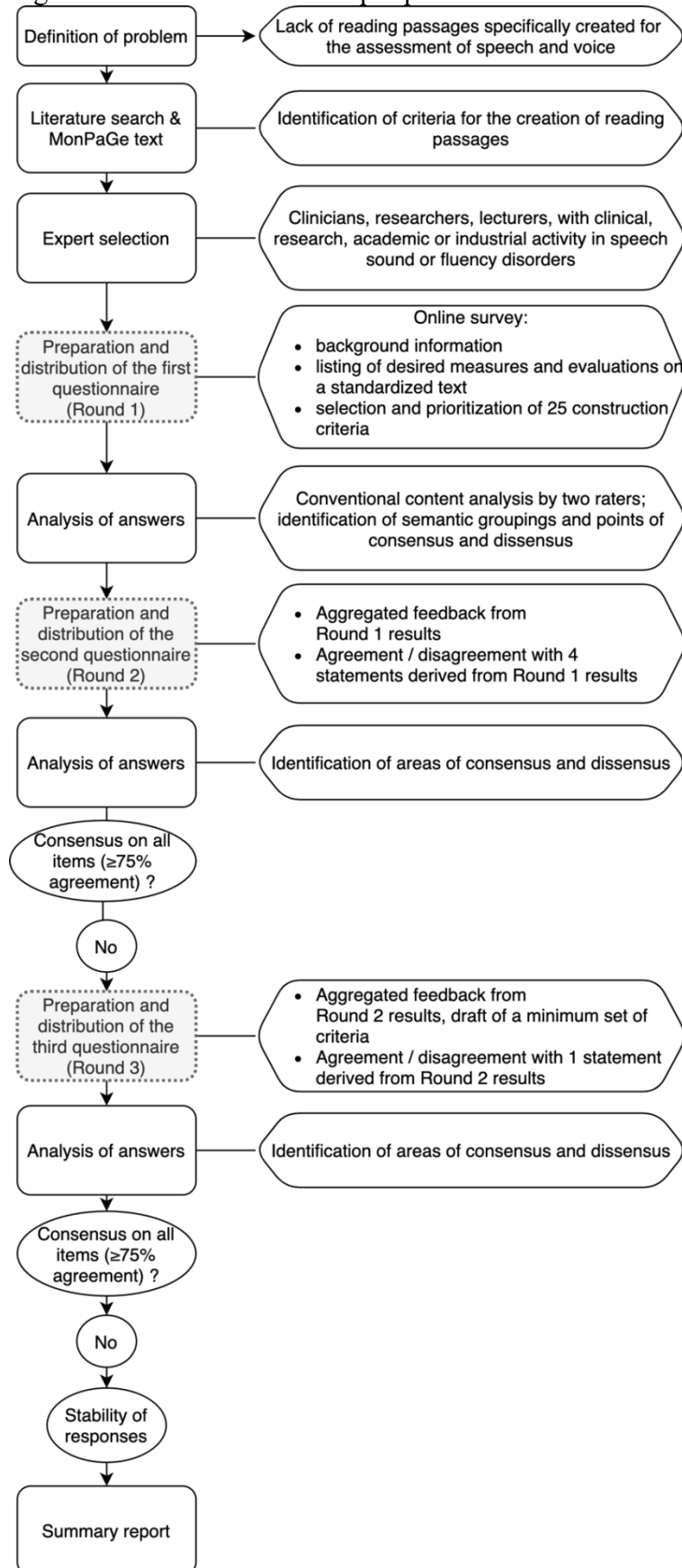


Figure 2. Mean ranks (MRs) for each of the criteria for the creation of a reading passage; black: most and best ranked (MRR <7.5 and rated by at least 50% of participants); dark grey: the least and the lowest ranked criteria (MRR >10 and rated by less than 25% of participants); hatched: adverse result between number of occurrences and MRR.

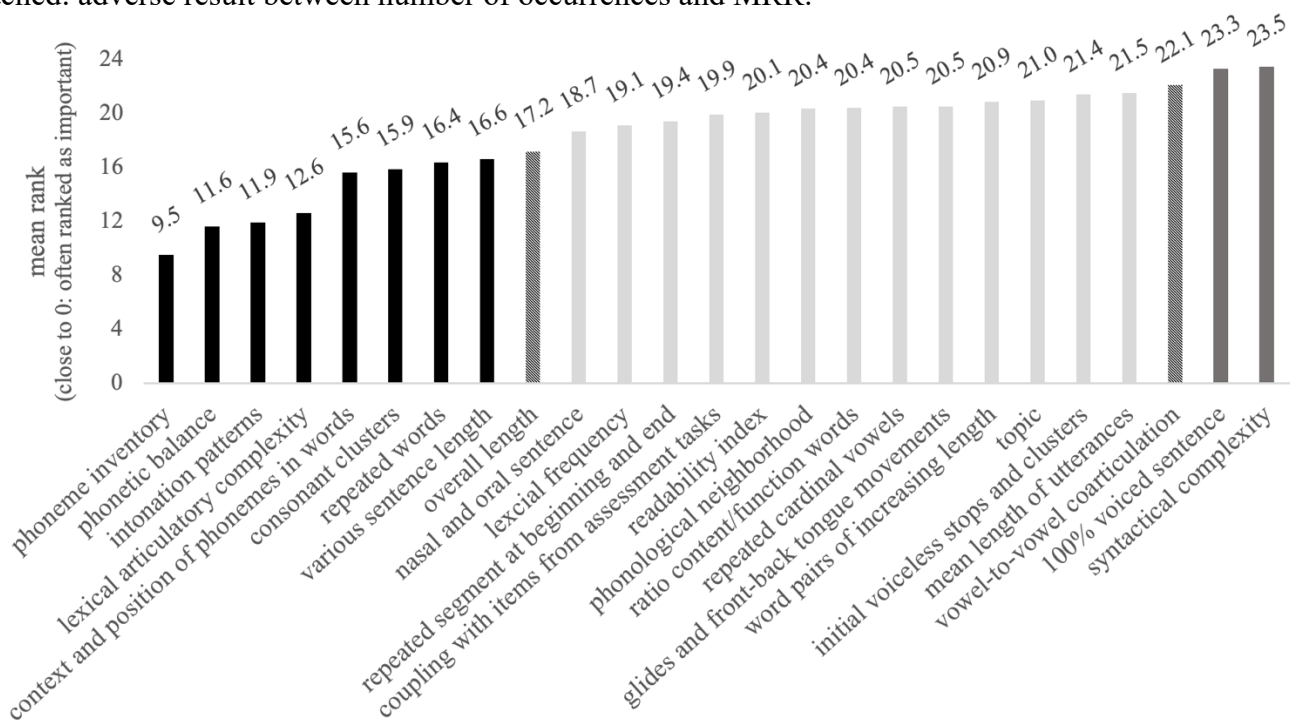


Figure 3. Minimal set of criteria to take into account when creating a standard reading passage for the assessment of speech and voice.

*Phoneme-level criteria:*

- I. Complete phonemic inventory (i.e. the text includes all of the target language's phonemes at least once);
- II. Phonetic balance (i.e. the included phonemes occur at approximately the same frequency at which they occur in ordinary conversation in the target language);
- III. Inclusion of (bi- and triconsonantal) clusters;
- IV. Taking into account the positions of phonemes in the words (beginning/middle/end), as well as their phonemic context;
- V. Taking into account vowel-to-vowel coarticulation (e.g. word pair /p a p a / - / p a p i /);

*Word-level criteria:*

- VI. Control of the articulatory/phonetic complexity of words;
- VII. Repeated occurrence of some words (error consistency, apraxia vs. dysarthria);

*Sentence-level criteria:*

- VIII. Presence of various intonation patterns (questions, statements, contrastive emphasis...);
- IX. Inclusion of various sentence lengths (breathing support, fatigability...);

*Global-level criteria:*

- X. Control of the overall length of the reading passage.

## **Figure captions**

Figure 1. Flowchart of the Delphi process used in this study.

Figure 2. Mean ranks (MRs) for each of the criteria for the creation of a reading passage; black: most and best ranked (MRR <7.5 and rated by at least 50% of participants); dark grey: the least and the lowest ranked criteria (MRR >10 and rated by less than 25% of participants); hatched: adverse result between number of occurrences and MRR.

Figure 3. Minimal set of criteria to take into account when creating a standard reading passage for the assessment of speech and voice.