

# Assessment of intelligibility in children

Tove B. Lagerberg

Division of Speech and Language Pathology  
Institute of Neuroscience and Physiology  
Sahlgrenska Academy at University of Gothenburg



UNIVERSITY OF GOTHENBURG

Gothenburg 2013

Cover illustration: Kommunikation by Ulf Lagerberg

Assessment of intelligibility in children

© Tove B. Lagerberg 2013

Tove.lagerberg@neuro.gu.se

ISBN 978-91-628-8851-0

Printed in Gothenburg, Sweden 2013

Ineko

*For Simon, Marvin and Hannes. Without the joy and support  
you give me every day, this would never have been possible.*

‘Låt inte känslorna stanna i bröstet. Prata. Bråka aldrig om  
pengar. Våga säga nej. Våga säga ja. Paradiset kan vara en  
plats på jorden.’

Kristian Gidlund



# Assessment of intelligibility in children

Tove B. Lagerberg

Division of Speech and Language Pathology, Institute of Neuroscience and Physiology,  
Sahlgrenska Academy at University of Gothenburg, Göteborg, Sweden

## Abstract:

**Aim:** The overall aim of this thesis was to investigate different aspects of intelligibility in children and to develop reliable and valid methods for assessment.

**Method:** Initially, four assessment methods were studied: multiple-choice assessment and transcription of single words, transcription of sentences and transcription of spontaneous speech. Audio recordings of 74 ten-year-old children with isolated cleft palate and/or 22q11DS and 11 children with typical development were included. Validity was examined through comparison of results for the children with and without deviant speech and between 'good' and 'poor' readers. Thereafter, spontaneous speech and single words taken from the STI-CH test and repeated after a model, produced by ten children with speech-sound disorder (mean age: 6.0 years) and ten children with typical speech and language development (mean age: 5.9 years), were recorded and presented to twenty listeners. Validity was studied through an investigation of the difference in intelligibility scores between the two groups and the correlation between intelligibility scores and PCC (Percentage of Consonants Correct) scores. Inter- and intra-listener reliability was investigated in relation to all assessments included in the thesis. Finally, three conditions for listener transcription of spontaneous speech were examined: listening to each utterance once, twice and three times.

**Results:** Inter- and intra-listener reliability was satisfactory for all methods included in the thesis. A statistically significant difference between outcomes for the four assessment methods studied initially was found and validity was low for all three reading-based methods. The intelligibility scores obtained for spontaneous speech correlated with PCC scores and differed statistically significantly between the two groups, indicating high validity. Statistically significant differences in terms of intelligibility scores were found between the three conditions investigated: the intelligibility score increased with the number of repetitions. Scores on STI-CH correlated with PCC scores and with intelligibility scores obtained using spontaneous speech, and they differed statistically significantly between the two groups, thus further confirming the validity of the test.

**Conclusion:** The choice of speech material and listener task has a significant impact on results when assessing intelligibility. Reading is not a suitable elicitation technique for ten-year-olds. The assessment procedure for spontaneous speech developed as part of the thesis can be recommended for intelligibility assessment, especially if the mean across several listeners is used, but the number of times a speech material is repeated to listeners must be reported. Finally, the single-words test developed as part of the thesis (STI-CH) showed good validity and reliability for the participants included.

**Keywords:** intelligibility, speech sound disorder, children, speech disorder, assessment

**ISBN:** 978-91-628-8851-0

# SAMMANFATTNING PÅ SVENSKA

Barn som har svårigheter med att prata, exempelvis när det gäller att uttala olika språkljud (bokstäver), kan få problem med att göra sig förstådda. Förståelighet (intelligibility) är ett begrepp som används inom logopedi och betyder ”förmågan att förmedla ett budskap via talad kommunikation”. Det finns många olika skäl till att problem med talet uppstår t ex artikulatoriska svårigheter som har neurologiska orsaker (t ex cerebral pares), avvikande anatomiska förutsättningar (läpp-käk-gomspalt) och tal- och språkavvikelse utan någon känd orsak (så kallad ”speech sound disorder”). Följden av dessa barns talsvårigheter kan bli en reducerad eller nedsatt förståelighet. En viktig uppgift för logopeden är att genom talträning eller genom andra åtgärder, exempelvis hjälpmedel, underlätta kommunikationen för dessa barn. Då kommunikation innebär att överföra ett budskap är det viktigt att kunna bedöma hur framgångsrik behandlingen varit när det gäller att öka barnens möjlighet att göra sig förstådda. Bedömning av förståelighet är också viktigt i forskning avseende konsekvenser av olika syndrom/talstörningar samt studier för att utvärdera olika logopediska åtgärder. För att en bedömningsmetod skall kunna användas behöver vi veta att den mäter just det man vill mäta (validitet) samt att den mäter detta korrekt (reliabilitet). Denna avhandling syftar till att undersöka och utveckla bedömningsmetoder för hur effektiv den talade delen av kommunikationen (d.v.s. inte gester och ansiktsuttryck) är när det gäller att överföra ett budskap samt att pröva dessa metoders validitet och reliabilitet.

I studie I prövades fyra olika bedömningsmetoder. De fyra metoder där barnen läste visade sig ge utslag lika mycket på barnens läsförmåga som deras talavvikelse, vilket tyder på låg validitet, det vill säga att testet inte mäter förståelighet utan något annat (läsförmåga avseende testmaterialet). Slutsatsen blev att ett test som utvecklats specifikt för barn och som inte baserades på läsning behövdes.

Två metoder utvecklades och prövades, en som kan användas när barnen pratar fritt och en som baseras på eftersägning av enstaka ord (STI-CH). För bedömningen av enstaka ord utvecklades 1000 ordlistor på ett sådant sätt att de skulle ge liknande resultat för ett barn oavsett vilken lista som valdes. Tjugo 5-8 åringar spelades in, 10 barn med typisk tal- och språkutveckling och 10 barn med talavvikelse (speech sound disorder), när de talade fritt och när de upprepade de ord som testledaren läste upp från en av ordlistorna. Tjugo logopedstudenter bedömde inspelningarna. Resultaten visade att båda testmetoderna var tillförlitliga dvs de relaterade till barnens talstörning på ett

sådant sätt att man kunde dra slutsatsen att det just var förståelighet som mättes (validitet) samt att den mättes på ett korrekt sätt (reliabilitet).

I den tredje delstudien prövades om bedömningen av förståelighet påverkades av hur många gånger lyssnaren fick höra det talaren sa. Inspelningar från 12 av barnen från studie II användes. Resultaten visade att det var en liten men statistiskt påvisbar skillnad i resultat avseende förståelighet beroende på hur många gånger man fick höra materialet. Slutsatsen blev att det är viktigt att i forskning redovisa hur många gånger man fått lyssna men att det inte spelade så stor roll vilket antal man valde. I denna studie blev det också tydligt att olika lyssnare kan ge mycket olika resultat avseende förståelighet trots att man lyssnar på samma inspelning från samma barn. En konsekvens av detta blir att man bör använda samma lyssnare vid uppföljningar eller många lyssnare och använda ett medelvärde.

Sammanfattningsvis visade avhandlingen att det är möjligt att bedöma förståelighet hos barn på ett tillförlitligt och inte alltför tidskrävande sätt med de nya utvecklade metoderna – (STI-CH) och bedömningen av förståelighet i spontantal. Det är dock viktigt att metoderna undersöks vidare exempelvis med barn med andra typer av talstörningar samt med större grupper.





# LIST OF PAPERS

This thesis is based on the following studies, referred to in the text by their Roman numerals.

- I. Johannisson [now: Lagerberg], T. B., Lohmander, A., Persson, C. Assessing intelligibility by single words, sentences and spontaneous speech: a methodological study of the speech production of 10-year-olds, *Logopedics Phoniatics and Vocology*, 2013. E-pub ahead of print.
- II. Lagerberg, T. B., Åsberg, J., Hartelius, L., Persson, C. Assessment of intelligibility using children's spontaneous speech: methodological aspects, 2013, *International Journal of Language and Communication Disorders*. In press.
- III. Lagerberg, T. B., Åsberg Johnels, J., Hartelius, L., Persson, C. Effect of number of repetitions on listener transcriptions in assessment of speech intelligibility in children, 2013. Manuscript.
- IV. Lagerberg, T. B., Åsberg Johnels, J., Hartelius, L., Ahlman, A-K., Börjesson, A., Persson, C. Swedish Test of Intelligibility for Children (STI-CH) – Validity and reliability of a computer-mediated single-word intelligibility test for children, 2013. Manuscript.

# CONTENT

1	INTRODUCTION.....	1
2	BACKGROUND.....	2
2.1	Speech, language and communication .....	2
2.1.1	Speech production .....	2
2.1.2	Perception/understanding .....	3
2.2	Speech-production disorders.....	4
2.3	The concept of ‘intelligibility’ .....	5
2.4	Speech disorders and intelligibility.....	6
2.5	Consequences of reduced intelligibility .....	6
2.6	Assessment.....	7
2.6.1	Type of speech material .....	7
2.6.2	Elicitation technique.....	8
2.6.3	Transmission medium .....	9
2.6.4	Listener characteristics .....	10
2.6.5	Listener task .....	11
2.6.6	Assessment in children.....	12
2.6.7	Reliability and validity .....	15
3	AIM.....	17
4	MATERIALS AND METHODS .....	18
4.1	Participants.....	18
4.1.1	Speakers .....	18
4.1.2	Listeners .....	22
4.2	Ethical considerations .....	22
4.3	Procedures.....	23
4.3.1	Development of the Swedish Test of Intelligibility for Children (STI-CH).....	23
4.3.2	Intelligibility assessment .....	25

4.3.3	Listener tasks, calculation of intelligibility score and listener conditions.....	25
4.3.4	Validity assessment.....	26
4.4	Statistical analysis and strategies.....	27
5	RESULTS.....	30
5.1	Effect of speech material, elicitation technique and listener task.....	31
5.2	Reliability and validity of assessment based on spontaneous speech.....	33
5.3	Impact of the number of repetitions of the speech material.....	34
5.4	Reliability and validity of STI-CH.....	35
6	DISCUSSION.....	37
6.1	Effect of speech material and elicitation technique.....	38
6.2	Listener and listener task.....	40
6.3	Reliability and validity.....	43
6.4	Clinical and research implications.....	46
6.5	The concept of ‘intelligibility’.....	47
6.6	Limitations and future research.....	48
	ACKNOWLEDGEMENT.....	49
	REFERENCES.....	51
	APPENDIX.....	57

# ABBREVIATIONS

CLP	Cleft lip and palate
DFA	Discriminant function analysis
ICC	Intra-class correlation
PCC	Percentage of consonants correct
SSD	Speech sound disorder
STI-CH	Swedish test of intelligibility- for children
SWINT	Swedish intelligibility test
TD	Typical speech and language development
VPI	Velopharyngeal impairment

# 1 INTRODUCTION

*'Because the fundamental purpose of speech communication is to be understood, intelligibility is the functional common denominator of verbal behavior.'* (Kent, Miolo & Bloedel, 1994, p. 81).

Intelligibility refers to how much a listener perceives of what a speaker is saying. It is often assessed using scales where the listener chooses among ratings ranging from, say, 'not at all intelligible' to 'completely intelligible', or by having a listener write down the words that he or she understands and then comparing this with a key script, calculating the percentage of correct words and using this as a measure of intelligibility. Having an intelligibility measure is often relevant to speech-language pathologists (SLPs) working with children and adults who have speech disorders that make them hard to understand. One important question that many, if not all, SLPs have asked themselves is whether the efforts they make really help the person with a speech disorder to make him- or herself understood more easily and, in the longer term, to use verbal communication in order to participate in social, work or educational activities.

A considerable amount of research has been conducted in this field. Its focus has differed somewhat depending on the type of speech problem involved and the age of the patients (adults or children), but broadly speaking most of the attention has been paid to the role of the listener in research concerning acquired speech disorders of neurological origin (dysarthria), whereas evaluation of intervention has been the predominant focus in studies of children, relating among other things to the effects of surgery on children with cleft lip and palate. Further, it might be claimed that the theoretical discussion of the concept of 'intelligibility' as such has been more in focus in some lines of research, while the pursuit of an effective and reliable method of assessment has been the main aim of other lines.

This thesis represents an attempt to link together some of these different aspects from the perspective of assessment methods for various types of speech disorders in children while also focusing on the concept of 'intelligibility' as such and on the role of the listener. The ultimate objective of the thesis is the development of an assessment tool for children, but findings from earlier research including both adults and children are discussed in order to provide a background.

## 2 BACKGROUND

### 2.1 Speech, language and communication

If we focus on the role of the speaker, the process of oral communication can be described as the path from having an idea of what we want to say, over formulating a message, to producing a chain of sounds (speech) that is audible and intelligible to a listener. The process thus involves a cognitive phase, a linguistic phase, a planning phase and a programming phase, where motor sequences have to be performed in a given order and during a specific time. Although these phases are tightly inter-related, the process is often divided into two aspects: language (cognitive and linguistic level) and speech (planning, programming and execution of motor activity). Speech production includes a sensory component as well, in that the system continuously receives feedback on the outcome of the process in the form of both auditory and tactile stimuli. Communication obviously consists of many additional components, such as facial expressions, gestures and other types of non-verbal communication, which play an important role in conveying a message to a communicative partner. However, the focus of the present thesis is on the speech signal and on the various consequences that a distortion in the production of speech may have.

#### 2.1.1 Speech production

Speech production is a complex motor activity that requires co-ordination of the respiratory, laryngeal and articulatory systems. Speech is first generated by air that is pushed from the lungs through the trachea, the larynx, the pharynx and the oral and nasal cavities. This expiratory air stream may cause the vocal chords in the larynx to vibrate, giving rise to voice (phonation). The air stream then passes through the pharynx and the oral and/or nasal cavities, where it is modified by the position and movement of the articulators (tongue, lips and palate), creating different speech sounds (Weismer, Yunusova & Bunton, 2012). For example, when the passage leading to the nasal cavity is closed by the elevation of the soft palate and the lips are closed and then released, this builds up intra-oral pressure resulting in a high-pressure sound, namely /b/. The sound /p/ is produced in the same way but without phonation, meaning that the result is a voiceless high-pressure sound. To create the nasal sound /m/, the lips are closed while the soft palate is kept open so that the air is forced up into the nasal cavity.

## 2.1.2 Perception/understanding

When it comes to the listener's role in oral communication, the process of perceiving a spoken message is not just a matter of capturing a sequence of consonants and vowels; rather, it is a question of drawing conclusions about the words intended by the speaker based on the whole picture, i.e. the overall sound environment (Miller, 2013). The information transmitted by the speech signal is of various types. One way to describe this is by referring to the segmental and suprasegmental levels. The segmental level includes individual speech sounds (phonemes). The suprasegmental one includes prosodic features such as stress, temporal aspects, intonation and word accent that become available to the listener when speech sounds are combined into syllables, words and phrases. Such prosodic features are particularly important when the information at the segmental level is less than optimal, and they are used by listeners to adapt and to use various speech-perception strategies to help them understand speech in a relevant way (Kent, 1992; Weismer & Martin, 1992). This may relate to deviant speech (see below), but also to non-deviant speech. For example, it may be difficult to understand a speaker with a foreign accent, even if he or she is at a high level in terms of syntax, grammar and pronunciation of individual speech sounds, if the prosody remains that of the person's first language.

To understand what has been said, however, the listener uses not only the speech signal but also knowledge about the context and the speaker as well as past experience. Context is of great importance for intelligibility in the sense that a word is easier to understand when it is presented within a sentence where the listener has access to both semantic (meaning) and grammatical clues than when it is presented alone. It should also be noted that the speech-perception system has a strong ability to adapt to different types of speech (such as foreign accents, hearing-impaired speech and dysarthric speech) through an experience-based process referred to as 'perceptual learning' (Borrie, McAuliffe, & Liss, 2012; Samuel & Kraljic, 2009).

The different stages involved in the understanding of speech have been widely discussed by scholars over the years. There are two main hypothetical processes: bottom-up and top-down, which may work in parallel. In the bottom-up process, understanding relies on the information contained in the acoustic speech signal, while the top-down process uses knowledge or anticipation about the content of what the speaker is saying (Garcia & Dagenais, 1998; Hustad & Beukelman, 2001; Kent, 1996; Lindblom, 1990). When the segmental, and perhaps also suprasegmental, features of speech are degraded, listeners may need linguistic cues to be able to use top-down

strategies in parallel with bottom-up processing to derive a meaning (Lindblom, 1990).

## 2.2 Speech–production disorders

Problems with speech production may be due to linguistic difficulties (language disorder), to articulatory difficulties (speech disorder) or to a combination of both. Phonological disorder is one example of a language disorder in children where the speech-sound system is incomplete. In the case of Swedish, the child may replace all velar sounds (e.g. /k/) with dental sounds (e.g. /t/) even though the child has no motor or structural problem in articulating /k/. A person with a speech disorder may retain intact linguistic abilities but lack the ability to use and co-ordinate the relevant motor processes, and the muscles and structures – e.g. the tongue, lips, jaw and palate – that are necessary to produce speech may be impaired or delayed.

There are many types of speech disorders with different aetiologies. A disturbance in the speech signal can occur for many reasons, for example structural impairments such as a cleft palate, which makes it impossible to close the passage between the oral and the nasal cavities. This is called a velopharyngeal impairment (VPI) and is common, for example, in children with cleft lip and palate (CLP) (Dzioba, Skarakis-Doyle, Doyle, Campbell, & Dykstra, 2013) or the chromosomal aberration called 22q11 deletion syndrome (Persson, Lohmander, Jonsson, Oskarsdottir, & Soderpalm, 2003). Another type of speech-motor disturbance, dysarthria, is the result of a neurological injury or condition, such as cerebral palsy (CP) (Hustad, Schueler, Schultz, & Duhadway, 2012).

In other words, there exist disturbances in at least three different stages of the process of speaking, caused by deviations in the linguistic system of speech sounds (e.g. phonological disorder), in the structure of the speech apparatus (e.g. VPI) or at the motor execution stage (e.g. dysarthria). However, regardless of the type of disturbance, the result can be problems in making oneself understood, i.e. reduced intelligibility. When it comes to speech-production disorders in children, the term ‘speech-sound disorder’ (SSD) has been used frequently in recent years (Allen, 2013; McLeod, Verdon, Bowen, & International Expert Panel on Multilingual Children's, 2013; Unicom, Hewat, Spencer, & Harrison, 2013). SSD is included as a diagnosis in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*Diagnostic and statistical manual of mental disorders : DSM-5*, 2013). It encompasses children with either phonological or articulatory disorder without any congenital or acquired medical or neurological condition. The



first diagnostic criterion for SSD is ‘Persistent difficulty with speech sound production that interferes with speech intelligibility [...]’ (*Diagnostic and statistical manual of mental disorders: DSM-5*, 2013, p. 44).

## 2.3 The concept of ‘intelligibility’

On a general level, it has been claimed that ‘Intelligibility is a *sine qua non* for successful spoken communication’ (Miller, 2013, p. 1; italics in original). In research and clinical contexts, the concept of ‘intelligibility’ has been defined and operationalised in a great many – sometimes radically different – ways. An early and widely used definition is the one proposed by Kent: ‘the degree to which the speaker’s intended message is recovered by the listener’ (Kent, Weismer, Kent, & Rosenbek, 1989, p. 483). Another widely used definition is ‘the degree to which the acoustic signal [...] is understood by the listener’ (Yorkston, Strand & Kennedy, 1996, p. 55). The first definition emphasises that a message, i.e. some kind of meaning, should be transmitted in an unspecified manner, while the second one is restricted to a single transmission channel, namely the acoustic signal (i.e. the speech signal), and thus only refers to the degree to which the speaker’s intended message is transmitted to the listener through that channel – without any contextual cues such as linguistic or visual cues from non-verbal communication (Yorkston et al., 1996).

A further concept used by dysarthria researchers which is closely related to intelligibility is ‘comprehensibility’, which refers to the ability to convey a message *in a communicative context*. This concept was introduced by Yorkston et al. (1996). However, Hustad (2008) argued that a more accurate term would be ‘contextual intelligibility’, emphasising the fact that this concept refers to what can be transmitted when the acoustic speech signal is not the only channel but is supported by visual cues (e.g. facial expressions and gestures) and contextual cues (e.g. knowledge of the topic). She further claimed that ‘comprehensibility’ (as indicated by the synonymous term ‘listener comprehension’) implies that the focus is more on the listener and on his or her ability to interpret the meaning of a message in a deeper sense, while the main focus of ‘intelligibility’ is on the lexical and phonetic accuracy of the speaker (Hustad, 2008). An additional concept, ‘functional intelligibility’, has recently been introduced by McLeod et al. (2012) to represent a speaker’s ability to convey a message *in daily life*. Similarly to ‘comprehensibility’, the aim of this new term is to shift the focus away from the speaker’s ability in order to obtain a more comprehensive picture of the ramifications of a speech disorder.

Nevertheless, the narrow definition of ‘intelligibility’ cited above (including only the acoustic speech signal) (Yorkston et al., 1996) is of value in situations where there is a need to isolate the characteristics of the speech signal, such as in the assessment of the efficacy of treatment in relation, for example, to articulatory training and compensatory strategies intended to improve the speech signal. One possible way of obtaining terminological clarity is to refer to ‘signal-dependent’ factors (relating to information perceived from the speech signal) and ‘signal-independent’ factors (relating to information from other sources, such as syntax, semantics and facial expressions) (Miller, 2013) – although it should be kept in mind that both signal-dependent and signal-independent factors play an important part in the process of transferring and understanding a message (Mattys, Davis, Bradlow, & Scott, 2012). Miller claims that much of the confusion about how best to assess and evaluate intelligibility derives from the failure to make this particular distinction.

## **2.4 Speech disorders and intelligibility**

Generally speaking, speech disorders are associated with reduced intelligibility, meaning that there is an obvious need to assess the intelligibility of people with speech disorders in order to evaluate the effectiveness of interventions made to help them improve their intelligibility. The relationship between specific speech or articulation problems and intelligibility has been investigated in a number of studies, especially in relation to adults. Listeners’ ability to understand what is said depends not only on perceiving the phonemes correctly at the segmental level, but is also affected by suprasegmental information such as prosody (Weismer & Martin, 1992). In general, speech deviations at the segmental level have been shown to exert a greater impact on intelligibility than suprasegmental deviations (Weismer & Martin, 1992). Tongue control in speech movement has been demonstrated to have a larger impact on intelligibility than lip and jaw control (Weismer et al., 2012), and there is a moderate correlation between articulation-test scores and intelligibility in children (Morris, Wilcox, & Schooling, 1995).

## **2.5 Consequences of reduced intelligibility**

The development of speech and language is an ongoing process during childhood. At the age of four, a child is generally expected to speak in a way that is fully intelligible to a listener (i.e. 100% intelligibility) (Coplan & Gleason, 1988 cited in Namasivayam et al., 2013), and it has been proposed

that if a four-year-old child has an intelligibility of less than approximately 60% (the percentage representing the proportion of words understood by an unfamiliar listener), speech therapy should be considered (Gordon-Brannan & Hodson, 2000). Intelligibility below this level may cause the child not to be understood by its peers or teachers, meaning that its ability to participate in social and learning activities will be reduced (Gordon-Brannan & Hodson, 2000). What is more, reduced intelligibility may have a negative influence on a child's thoughts and feelings about his or her ability as a communicator (and thus his or her attitude to communication) (Johannisson et al., 2009). In a longitudinal study by Havstam, Sandberg, & Lohmander (2010), the attitude to communication in ten-year-old children with cleft palate correlated statistically significantly with overall global measurements of intelligibility.

## 2.6 Assessment

*'Given the pivotal position of intelligibility in defining successful communication and therefore its centrality as an outcome measure in speech-language therapy, there is a definite place for routine objective assessment of intelligibility.'* (Miller, 2013, p. 1). Auditory-perceptual judgements are an essential but challenging component in the field of speech-language pathology. The various considerations that need to be taken into account in the performance of this task are described in an article by Kent with the telling name of 'Hearing and believing' (1996). In the case of intelligibility, there is also a need to reckon with the variation in ways of defining the concept, as described above, which entails that there are a number of choices to be made when it comes to assessing the level of intelligibility. As is clear from the above discussion about the concept of 'intelligibility', the ability to convey a message is only partially dependent on the speaker. Other potentially important factors in this process include the type of speech material used for the assessment, the elicitation technique used by the examiner, the transmission medium, the listener's characteristics and the task to be carried out by the listener. Arguably, all of these factors must be considered in the assessment of intelligibility.

### 2.6.1 Type of speech material

The type of speech material may play a role, both for the speaker's ability to produce the speech (single words may be easier to produce than longer utterances) and for the amount of contextual information that is provided to the listener (single words give less information than longer utterances). One example of a practical implication of this is that speakers with severe dysarthria are generally less intelligible in sentences than in single words,

while the opposite is true of speakers with mild dysarthria (Lillvik, Allemark, Karlström, & Hartelius, 1999; Yorkston & Beukelman, 1978). In fact, the contextual information provided in continuous speech such as sentences seems to be the most helpful to listeners in the middle range of the continuum from unintelligible to intelligible speech (Miller, 2013; Sitler, Schiavetti, & Metz, 1983).

The use of a structured speech material such as a list of predetermined single words or sentences is associated with both advantages and disadvantages. The advantages include that this makes it possible to control the identity and frequency of the phonemes included as well as the level of articulatory complexity. It is also easier to calculate the intelligibility score (i.e. the percentage of correctly perceived words or syllables) based on the listeners' answers about what they perceived, since it is known with certainty what the speaker intended to say. One disadvantage is that a listener who repeatedly uses such a material to make assessments will soon know what words or sentences are included. To prevent this, it is necessary to create a sufficiently large pool of words or sentences from which speech material can be drawn. This may be especially important in clinical contexts, where the number of listeners available (e.g. the SLPs working at a certain clinic) is often restricted. A further disadvantage is that a structured speech material may lack ecological validity, i.e. it may not be representative of the speaker's speech in daily life, for example because many of the words included are not part of the speaker's active (or even passive) vocabulary.

From the perspective of ecological validity, spontaneous speech may be better suited than a predetermined material as the basis for an assessment (which is, after all, typically intended to closely reflect the speaker's performance in daily life). However, this type of speech material has some major drawbacks: first, there is no way to be certain about what the speaker intends to say; and second, the speaker is able to avoid words or phonemes that he or she finds difficult to produce. When a speaker's intelligibility is severely reduced, an additional problem is how to identify the denominator for the calculation of the intelligibility score, i.e. how to determine the total number of words in a speech sample (Flipsen, 2006).

## **2.6.2 Elicitation technique**

Speech can be elicited by having a speaker read out loud, name pictures orally or repeat after a model, or by asking open questions in a conversational context. In the case of adults, reading is a frequently used elicitation technique for intelligibility assessment (Lillvik et al., 1999; Yorkston &

Beukelman, 1981). Yorkston claims that the choice of elicitation technique should be based on the speaker's ability, even though reading is preferable to repeating after a model since the latter technique may yield higher intelligibility scores in adults with dysarthria (Yorkston & Beukelman, 1981).

One major advantage of letting the speaker read the words, sentences or text out loud is that the intended target words are known. However, this may be true only of individuals with well-developed reading skills and so may not apply to children, where difficulties in reading (accuracy and/or fluency) rather than in speech production may constrain an individual's performance. What is more, there is research suggesting that phonological difficulties are an underlying source common to both speech-sound deficits and reading difficulties (Pennington & Bishop, 2009). This means that it could be difficult to distinguish speech problems from reading difficulties in children if reading is used as an elicitation technique.

Picture-naming avoids the problem of reading skill as a confounder and also probably yields articulatory behaviour which is closer to that found for free speech, but on the other hand there is even less certainty as to whether the speaker tries to produce the word intended by the test designer, since the speaker may interpret the picture differently. In addition, finding a sufficient number of relevant pictures can be a major challenge.

The option of repeating after a model has been questioned because of the articulatory help it involves, especially if the speaker sees the model producing the word (Kwiatkowski & Shriberg, 1992). One way to mitigate this is to use recordings of a model instead of a live model.

Finally, as already mentioned, the method of asking open questions to generate spontaneous speech has high ecological validity but makes it more difficult to know what the speaker intended to say and gives him or her the unfortunate opportunity to avoid 'difficult' words.

### **2.6.3 Transmission medium**

A material can be presented to the listener in two ways: audio only or audiovisual. The choice may be affected by the operationalisation of the concept of 'intelligibility': additional non-verbal information, such as visual information that the listener receives if he or she sees the speaker, should not be included if the definition used is that of Yorkston et al. (1996), which restricts the concept to the speech signal, meaning that audio only should then be used. Generally, however, although visual cues provide the listener with additional information, for example about place of articulation and facial

expressions, it is not certain that the audiovisual mode of transfer gives higher intelligibility. Research in this area provides no clear answers (Hustad & Cahill, 2003). Even though studies have shown that the severity of the speech disorder and the presence of motor impairment may play a role (Hustad & Cahill, 2003), it is not clear under what circumstances and for what purposes audio-only or audiovisual presentation, respectively, is preferable.

## 2.6.4 Listener characteristics

Various listener characteristics such as age, sex and familiarity with the speaker and with his or her dialect or speech disorder have been investigated in several studies. In a study by Pennington and Miller (2007), age and familiarity with the speaker's dialect were found to have no effect on intelligibility scores. McHenry (2011) found no difference between the intelligibility scores obtained by various listeners based on age, sex or level of education.

A further variable (or set of variables) related to the listener that has been discussed and investigated is 'familiarisation' with a specific speaker or the features of a specific speech disorder (Hustad & Cahill, 2003; Tjaden & Liss, 1995). This is based on the assumption that a listener is able to interpret a speech signal more accurately if he or she has previously been exposed to that signal, or a similar one. There is a consensus to some extent that listener familiarisation does yield higher intelligibility scores (Hustad & Cahill, 2003), but the variability of the speech disorder may exert an influence: if the speech deviances are irregular and unpredictable, this effect is not certain to occur. Further, the severity of the speech disorder also seems to influence the effect of familiarisation, at least for speakers with severe dysarthria (Hustad & Cahill, 2003).

Another type of familiarisation involves the listener being aware – to a varying extent – of the content of the speech material; this may influence performance on word-recognition tasks since it, so to speak, reduces the number of possible options to choose from. For example, a listener who has children who are of the same age as a speaker may be more in the habit of hearing the words used. One way to control for this effect when using a predetermined speech material is to let the listeners read all possible words included in the test beforehand (Hodge & Gotzke, 2007), in order to make all of them equally familiar with the words.

## 2.6.5 Listener task

The most frequently used method of intelligibility assessment involves spontaneous speech being evaluated by a listener using a scaling procedure where the listener is asked to award a grade on a scale where the end points are, say, ‘always intelligible’ and ‘completely unintelligible’ (Whitehill, 2002). However, Whitehill (2002) has claimed that the validity and reliability of this method have not been sufficiently evaluated. As regards validity, one question that needs to be asked is whether the raters are able to distinguish intelligibility from the severity of the speech disorder or from ‘acceptability’ – i.e. how deviant or strange the speech sounds to the listener (Whitehill, 2002).

In addition, Schiavetti has argued (1992) that since there are methods available to measure intelligibility at the ‘ratio level’, they should be used. The ratio level is the highest level of measurement and thus higher than the ‘ordinal level’ to which scaling methods belong. Measures on the ratio level are commonly obtained by means of a word-recognition task, where the percentage of words correctly understood by a listener is determined and commonly referred to as the ‘intelligibility score’. The words may be single words or may be part of sentences or spontaneous speech, and the listener task may be multiple-choice (closed-set) response or transcription.

Such intelligibility scores based on the same speech material typically vary depending on the listener task. For single words, multiple-choice response (where the listener has a number of related responses to choose from) has been shown to give less variable (McHenry, 2011) and higher (Yorkston & Beukelman, 1978) intelligibility scores than transcription. As regards adults with dysarthria, multiple-choice has been recommended for severe speech disorders or to detect subtle changes over time while transcription can be used to make comparisons with typical speech or for mild-to-moderate speech problems (Yorkston & Beukelman, 1981). A carefully designed multiple-choice task can also provide a basis for qualitative analysis to identify the deviances in the speech signal that impair intelligibility. The ‘gold standard’ procedure for the assessment of intelligibility based on spontaneous speech is generally considered to be as follows (Gordon-Brannan & Hodson, 2000; Hodge & Gotzke, 2007): a transcription of speech made by a listener with access only to the speech signal – i.e. no contextual cues such as visual information or knowledge about the speaker or the context – is compared with a key script (representing the ‘correct’ transcription) made by caregivers and clinicians using all available clues such

as visual and contextual information. The intelligibility score is the percentage of words correctly understood by the listener.

## **2.6.6 Assessment in children**

Assessment in children involves some specific challenges that are not present in the assessment of adults, and there is a severe lack of knowledge about the appropriate approaches to assessment in research and clinical practice (Kent et al., 1994; Miller, 2013). For instance, children have a smaller lexicon, which makes it harder to use a large pool of words. What is more, the option of having speakers read the material out loud, which is frequently used with adults, is not available at all for pre-school children and may be less reliable for school-age children. Finally, in the case of spontaneous speech, it may be difficult to elicit material from children because they are more likely than adults to be shy or unsure of their ability to speak (making it difficult to obtain a sufficiently large speech sample).

As regards structured speech materials (single words or sentences), several tests for different age groups or speech disorders, using different elicitation techniques and listener tasks, have been described in the literature. Table 1 gives an overview of the characteristics of some of those most often referred to, including information about the methods used to investigate their validity (an aspect of tests discussed in the next section).

There thus exist a number of tests to assess intelligibility in children, but there are none for Swedish-speakers. The development of such a test was therefore one aim of the work underlying the present thesis.



*Table 1. Speech-intelligibility tests for children including single words*

	Speech material	Target group	Elicitation	Listener task	Validity examined using
Weiss intelligibility test (Weiss, 1982)	Single word & spontaneous speech	Children & adolescents	Picture-naming	Transcription	Correlation with overall impressions of intelligibility
Preschool Speech Intelligibility Measure – PSIM (Morris et al., 1995)	From Assessment of intelligibility of dysarthric speech (Yorkston & Beukelman, 1981)	Pre-school children	Repetition after examiner and picture-naming	Multiple-choice	Goldman-Fristoe Test of Articulation & teacher intelligibility ratings
Children’s Speech Intelligibility Measure – CSIM (Morris & Wilcox, 1999)	Single word	School-age children	Repetition after examiner	Multiple-choice	Goldman-Fristoe Test of Articulation & ratings of intelligibility by clinician
Speech Intelligibility Probe for Children with Cleft Palate – SIP-CCLP (M. Hodge & C. L. Gotzke, 2007)	Single word	Children with CLP, 3–6 years old	Repetition after recordings	Multiple-choice or transcription	Spontaneous speech and TD vs. CLP
Computer-mediated single-word intelligibility test (Zajac, Plante, Lloyd, & Haley, 2010)	Monosyllabic words	Children with CLP	Repetition after examiner	Transcription	PCC in single words



## 2.6.7 Reliability and validity

For an assessment method to be useful, it must measure the variable of interest rather than something else (i.e. it must have high validity). The method must also measure that variable with high precision (i.e. it must have high reliability). In studies of intelligibility, reliability is often investigated by comparing results from different listeners (inter-listener reliability) and results from repeated assessments by the same listener (intra-listener reliability). Inter-listener reliability is often analysed by means of intra-class correlation (ICC). Hodge and Gotzke (2007) reported ICCs of 0.99 for children with cleft palate and 0.86 for children without cleft palate (in both cases indicating excellent reliability) for a task involving transcription of spontaneous speech. Another study, where assessment was based on orthographic transcription of word lists, had ICCs ranging from 0.94 to 0.99 (excellent reliability) for different listener groups (Zajac et al., 2010). Intra-listener reliability is often reported in terms of correlations or point-by-point agreement. Rates of point-by-point agreement reported in various studies generally range from 75% to 92% but are sometimes lower for speakers with more distorted speech: as far down as 58% (Hodge & Gotzke, 2007). Correlations are often very high, in the ranges of Pearson product-moment correlation coefficient  $r = 0.92-1.00$  (Gordon-Brannan & Hodson, 2000; Zajac et al., 2010) and Spearman's rank correlation  $\rho = 0.82-1.00$  (Lillvik et al., 1999).

Validity in intelligibility assessment is commonly studied by comparing results for groups with and without distorted speech (Hodge & Gotzke, 2007; Zajac et al., 2010) and by correlating results with related variables such as scores on articulation tests or intelligibility scores obtained on the basis of other speech materials (Lillvik et al., 1999; Morris et al., 1995; Zajac et al., 2010). Another possibility is to compare the results with those obtained using another test which is known to measure the same variable, but since the reason for creating a new test is often the lack of existing reliable methods, this may be problematic (Streiner & Norman, 2008).

That the use of different assessment methods has an impact on the intelligibility scores obtained has been known for a long time. This was shown as early as 1978 by Yorkston and Beukelman. Since that time, the importance of using valid and reliable assessment methods in clinical work and research (and the importance of discussing methodological issues thoroughly) has been pointed out in several reviews (e.g. Kent et al., 1994; Whitehill, 2002). In spite of this, the concept of 'intelligibility' is still sometimes used without sufficiently careful consideration being given to its

content or implications. Methods found to be less reliable, such as rating scales and overall estimations, remain widely used in clinical work and in research on the effects of interventions (Miller, 2013). There is thus clearly a need to develop methods for the assessment of intelligibility that are reliable and valid. In this connection, there is also a need for a discussion of the concept of ‘intelligibility’ as such.

### 3 AIM

The overall aim of the studies on which this thesis is based was to investigate different aspects of intelligibility in children and to develop reliable and valid methods for assessment.

The specific aims of each of these four studies were to investigate:

- I. how the choice of speech material, elicitation technique and listener task affects intelligibility scores in ten-year-olds with and without deviant speech;
- II. the reliability and validity of a method for assessing intelligibility based on the orthographic transcription of words perceived as intelligible in spontaneous speech;
- III. the impact of the number of repetitions of the speech material on listener transcriptions in the assessment of intelligibility;
- IV. the validity and reliability of the Swedish Test of Intelligibility for Children (STI-CH), which is based on single words.

## 4 MATERIALS AND METHODS

A total of 131 participants were included in the studies underpinning this thesis: 105 speakers and 26 listeners. Study I included speakers from an outcome study where four listeners were invited to participate. For the three remaining studies, new speakers and listeners were recruited (see Table 2). Different types of speech materials were collected for the four studies to be used as the basis for the calculation of intelligibility scores. These speech materials were single words read out loud (Study I), single words repeated after a model (Study IV), sentences read out loud (Study I) and spontaneous speech (Studies I–IV). In addition, recordings of picture-naming tests were used to calculate the Percentage of Consonants Correct (PCC) (Studies II and IV).

### 4.1 Participants

#### 4.1.1 Speakers

Study I included 74 ten-year-old children with isolated cleft palate and/or 22q11 deletion syndrome who had been assessed, in an outcome study, with respect to speech function (compensatory articulation and velopharyngeal impairment). The speech assessment, made by SLPs with experience in the field of cleft palate using a speech material designed for the assessment of speech in children with cleft palate, showed that 25 of these children were judged to have deviant speech (the ‘Clinic+ group’) while 49 of them were not (the ‘Clinic group’). A further eleven children with typical development participated in the study as a comparison group, labelled ‘Controls’.

For the following studies (Studies II, III and IV), it was decided to include a group of speakers representing a wider range of speech difficulties in order to avoid ceiling effects and to investigate the impact of this variability on intelligibility scores using the various methods chosen. In addition, there was an aim to include younger age groups since one objective was to develop test methods for children younger than ten.

*Table 2. Overview of participants, materials and procedures in the four studies*

Speakers	Listeners	Material	Listener task	Aim
Study I 74 ten-year-olds with isolated cleft palate and/or 22q11DS 11 ten-year-olds with typical development	2 SLPs (intelligibility) 2 SLPs (reading)	68 words and 10 sentences (Swedish Intelligibility Test – SWINT) Audio-tape recorded spontaneous speech	Orthographic transcription Multiple-choice Rating of reading skills	Compare the impact of different speech materials on intelligibility scores Investigate whether SWINT is suitable for children
Study II 10 children with speech-sound disorder (average age 6.0 years) 10 children with typical speech and language development (average age 5.9 years)	20 SLP students (intelligibility) 2 SLP students (PCC)	Spontaneous speech 76 words, SVANTE (PCC)	Orthographic transcription	Investigate the validity and reliability of the new suggested method for the assessment of intelligibility based on spontaneous speech
Study III 12 of the children from Study II with PCC score < 90%	Same 20 SLP students as in Study II	Spontaneous speech	Orthographic transcription	Investigate the impact on intelligibility scores of the number of times the speech material is repeated to the listeners
Study IV The same children as in Study II	Same as in Studies II and III (but 2 students were replaced by recent SLP graduates)	Swedish Test of Intelligibility for Children – STI-CH (60 x 2 words) Spontaneous speech (same as in Study II) SVANTE (PCC) (same as in Study II)	Orthographic transcription	Investigate the validity and reliability of STI-CH





In Studies II and IV, a total of twenty children participated as speakers. Ten children with speech-sound disorder (the ‘SSD group’) were recruited from the Department of Paediatric Speech and Language Pathology at the Queen Silvia Children’s Hospital in Gothenburg. All of these children had been diagnosed as having a speech and language disorder that affected intelligibility according to the treating SLP (age range: 4:6–8:3 years;  $M = 6.0$  years;  $SD = 1.0$ ). In addition, ten children with typical speech and language development (the ‘TD group’) were recruited through contacts with schools and pre-schools in the same area; the exclusion criterion for these children was past or present contact with an SLP (age range: 4:8–7:4 years;  $M = 5.9$ ;  $SD = 1.1$ ). All twenty children had normal hearing and Swedish as their strongest language, as reported by their parents. Finally, twelve of these children were chosen as participants for Study III. This was because, to avoid ceiling effects, only those children whose PCC score was below 90% were included (Table 3).

*Table 3. Age (year:month), sex (F = female, M = male) and PCC (percentage of consonants correct) score of speakers in Studies II–IV. SSD = children with speech-sound disorder; TD = children with typical speech and language development. The ID numbers assigned in Study III to the children who participated in that study are given in parentheses*

SSD group				TD group			
Participant	Age	Sex	PCC	Participant	Age	Sex	PCC
SSD-1 (S1)	4:6	M	59	TD-1 (S11)	4:8	M	79
SSD-2 (S2)	5:6	F	60	TD-2	4:11	F	98
SSD-3 (S3)	6:7	M	54	TD-3	7:3	M	99
SSD-4 (S4)	5:11	F	61	TD-4 (S12)	5:0	F	72
SSD-5 (S5)	5:11	M	70	TD-5	5:3	M	97
SSD-6 (S6)	8:3	M	81	TD-6	7:4	M	100
SSD-7 (S7)	6:6	F	65	TD-7	7:3	F	100
SSD-8 (S8)	6:6	F	49	TD-8	6:6	F	100
SSD-9 (S9)	5:2	M	61	TD-9	5:10	M	96
SSD-10 (S10)	5:4	F	61	TD-10	4:10	F	100

## 4.1.2 Listeners

In Study I, an attempt was made to have a similar design as in many intervention-outcome studies, where the number of listeners is often limited (Whitehill, 2002). Two SLPs with experience in the field of cleft palate served as listeners for the intelligibility assessment while two other SLPs assessed whether the children had difficulty reading the words and sentences. By contrast, the design of the subsequent three studies placed more emphasis on the role of the listeners, meaning that a larger listener group was recruited. The intention was to use the same listeners in Studies II, III and IV. Twenty SLP students served as listeners in Study IV, but two of them were not able to participate in Studies II and III and were replaced with two recent graduates from the SLP study programme. The assessment of PCC in Studies II and IV was made by two SLP students (who did not serve as listeners). All listeners in Studies II–IV were female and between 20 and 35 years old. They all had normal hearing and Swedish as their strongest language, according to self-reports.

## 4.2 Ethical considerations

The parents of the children participating in the studies had been informed about the nature of the study before agreeing to their participation and had signed an informed-consent form. The children were also given brief information about the study before they agreed to participate. Ethical approval was obtained from the Ethics Committee of the Medical Faculty of the University of Gothenburg for Study I and from the Regional Ethical Review Board in Gothenburg for Studies II–IV.

In retrospect, it might have been useful to collect more information about both the children and the listeners participating in the studies, for example about the children's general development or about the listeners' hearing ability. However, this would have required a supplementary application for ethical approval. What is more, there is an ethical balance to be struck as regards how much information should be collected about the participants: their right to privacy must be respected, and participation in the study must not involve excessive effort. Against that background, the amount of information gathered here was considered to be reasonable given the purpose of the study. The children in Study I received a cinema ticket as compensation; otherwise none of the participants was given any compensation. The SLP students gained an insight into the research process, which may have been a factor motivating them to participate.

## 4.3 Procedures

### 4.3.1 Development of the Swedish Test of Intelligibility for Children (STI-CH)

A major objective of the work underpinning the present thesis was the development of the Swedish Test of Intelligibility for Children (STI-CH). The development of this test is described in detail in Paper IV, where its validity and reliability are also investigated. In brief, there were two main aims for STI-CH: (1) the words included should be ones used by children in their daily life; and (2) each word list should be representative of children's speech in terms of the frequency of various phonemes and consonant clusters and in terms of word length. The word lists were drawn from a word bank (containing 1389 words) that had been built in a previous study (Case, Forsberg, & Uppman, 2009) using audio recordings of children's speech during play in an after-school recreation centre. To create STI-CH, all homonyms, all words that could be perceived as offensive and all words that were not real were excluded. This resulted in a word bank containing 1243 words. To ensure, as far as possible, that the lists would be representative of the children's level of articulation in daily life and that the individual lists would be equally difficult to produce for the children, a number of actions were taken. First, the words were tagged manually with respect to their initial, medial and final phonemes as well as the number and type(s) of consonant clusters and the number of syllables. Then software created specially for the study was used to compile word lists, based on these tags, according to three selection rules requiring each individual word list to include the same proportions as the overall word bank of:

- certain phonemes in initial, medial and final position;
- certain types of consonant clusters; and
- words with certain numbers of syllables.

The first rule covered only phonemes with a prevalence of more than 2 per cent in a certain position. Once all requisite phonemes in the requisite positions were included, the remaining words (up to 60) in a list were randomly selected as regards individual phonemes. This resulted in word lists that were specified with regard to phonemes, clusters and word length (Tables 4 and 5). A total of 1000 different word lists consisting of 60 words each were created (for examples of word lists, see Appendix 1).

*Table 4. Overview of the composition of word lists as regards consonant clusters and word length*

Consonant clusters:	none	23
	two consonants (not including /s/ or /r/)	16
	three consonants (not including /s/ or /r/)	4
	consonant clusters including /s/ or /r/	17
	Word length:	
one syllable	16	
two syllables	32	
three syllables	9	
four syllables	3	

*Table 5. Minimum number of instances of phonemes by word position in the STI-CH word lists. Note that a word will typically register in several places in the table. For instance, pall /pal/ 'stool' scores one instance each for initial /p/, medial /a/ and final /l/*

		Initial	Medial	Final
Plosives	/p/	2		
	/b/	3		
	/t/	2	5	5
	/d/	2	3	1
	/k/	2	3	1
	/g/	1	3	2
Fricatives	/f/	4		
	/v/	2		
	/s/	4	2	2
	/ʃ/	1		
	/z/	1		
	/j/	3		
	/h/	4		
Nasals	/m/	3	2	
	/n/	1	3	7
	/ŋ/	-	1	
	/r/	2	4	10
	/ʁ/ /dʁ/	-	2	
	/l/	4	4	2
Vowels	unrounded vowels	3	31	4
	rounded vowels	2	25	2
	/a/	2	20	12

## 4.3.2 Intelligibility assessment

### Speech material

Three types of material were used to assess intelligibility in the various studies: the Swedish Intelligibility Test (SWINT) (Lillvik et al., 1999), the Swedish Test of Intelligibility for Children (STI-CH) and spontaneous speech.

SWINT, which was used in Study I, was developed for adults with dysarthria and is based on earlier tests using minimal sets, such as the Assessment of Intelligibility of Dysarthric Speech (Yorkston & Beukelman, 1981). It includes lists of 68 randomly selected words and ten nonsensical sentences each including four randomly selected content words. To minimise contextual cues, these sentences are grammatically correct but semantically impossible, e.g. ‘En arg dröm skakar en ek’ (An angry dream is shaking an oak). The recordings of the speakers in Study I reading the words and sentences from SWINT were made on the same occasion as the recordings for the assessment of deviant speech (see above). STI-CH was used in Study IV to assess single-word intelligibility.

Spontaneous speech was used in all four studies. The material was collected in connection with the single-word testing when the children spoke freely about school or what they did in their free time. The examiner’s utterances were removed from the recordings and the spontaneous speech produced by the children was divided into utterances of 1–18 words (utterances consisting of only ‘yes’ or ‘no’ were not included). For each child, a speech sample consisting of utterances totalling approximately 100 words was prepared. In Study I, speech material from children who produced at least 50 words was included. In Studies II and IV, there were two cases where the original recording contained fewer than 100 words (SSD-3: 50, TD-1: 52). Finally, eight children in the SSD group produced enough speech for two different samples to be created from their output. This made it possible to compare results for different speech samples from the same child.

## 4.3.3 Listener tasks, calculation of intelligibility score and listener conditions

Four different listener tasks were used:

- (1) orthographic transcription of single words (Studies I and IV);

- (2) a multiple-choice task for single words, where listeners were to choose one out of five words presented to them (Study I);
- (3) orthographic transcription of sentences (Study I);
- (4) orthographic transcription of spontaneous speech (Studies I–IV).

For single words and sentences, the percentage of the words that the listeners had understood correctly was used to calculate the intelligibility score. As regards spontaneous speech, the intelligibility score was calculated in two different ways. In Study I, the listeners made their transcriptions in handwriting and were instructed to write down the words that they understood. If they were uncertain about a word, they were told to draw a circle around it; and if a word was not understandable, they were to mark its place in the utterance with a cross. This yielded three different categories: (1) words understood (i.e. perceived as understood); (2) words guessed; and (3) words not understood. By contrast, in Studies II–IV a suggested improvement to that method was tested. Computer software was used both for transcription and for the calculation of the intelligibility score, and a further change was the use of syllables instead of words as the base unit, in an attempt to avoid the problem of word counting in unintelligible sound strings. Hence, the listeners were instructed to transcribe orthographically, using the keyboard and the software, all words that they understood, and to mark each syllable which they did not understand with ‘0’. The intelligibility score for each speaker was calculated as follows:  $\text{Intelligibility} = \frac{\text{Total number of syllables in transcribed words (not including '0's)}}{\text{Total number of syllables (including '0's)}} \times 100$ .

Study III used three different listener conditions: the listeners first heard and transcribed an utterance once (C1). Then the same utterance was played a second time and the listeners were asked to modify their transcription if they felt this to be appropriate (C2). Finally, when they had listened to all utterances from one child in this manner, they listened to the utterances again, one at a time and in the same order, and were again asked to modify their transcription if appropriate (C3).

#### **4.3.4 Validity assessment**

To explore the validity of the intelligibility-assessment methods investigated in the studies underpinning this thesis, the relationship between the intelligibility scores obtained and two other characteristics of the children was analysed. Specifically, it was assumed that intelligibility is related to the percentage of consonants correct (PCC) (McLeod, Harrison, & McCormack,

2012; Shriberg & Kwiatkowski, 1982; Zajac et al., 2010) but that it is not related to reading ability.

### **Percentage of consonants correct (PCC)**

The PCC metric was created to assess the severity of involvement, including intelligibility (Shriberg & Kwiatkowski, 1982), and it has been widely used in research concerning children with speech and language disorders (Brosseau-Lapre & Rvachew, 2013; Chen et al., 2010; Lundeborg & McAllister, 2007; McLeod, Harrison, McAllister, & McCormack, 2013; McLeod et al., 2012). Originally created to be used on spontaneous speech, it has later been used on single words as well (Klinto, Svensson, Elander, & Lohmander, 2013; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997a; Zajac et al., 2010). In this thesis, PCC was assessed on the basis of single words collected from a picture-naming task using SVANTE, a Swedish articulation test (Lohmander et al., 2005), on the same occasion as the recordings of STI-CH and the spontaneous speech were made. The assessment was performed by the two SLP students who made the recordings, as a consensus assessment in accordance with the scoring rules of Shriberg and Kwiatkowski (1982).

### **Reading ability**

In Study I, the validity of SWINT was investigated using reading ability as measured on the basis of its 68 single words and 10 sentences. Two SLPs who were not involved in the listening task to assess intelligibility listened to all 85 children's sentences and words, indicating whether each child 'read without big problems, making only a few mistakes' or 'read in a hesitant way, making many mistakes'. Children who were assessed as belonging to the second category by both SLPs were considered 'poor readers' while the others were considered 'good readers'. (Note that this does not imply that these children were in fact poor readers in a general sense, only that they had difficulties reading the material in SWINT – which, having been designed for adults, was considered to be a possible confounder in the assessment of the speech intelligibility of children.)

## **4.4 Statistical analysis and strategies**

A detailed overview of the statistical methods applied is provided in Table 6. In Study I, the impact of four different assessment methods on intelligibility scores was compared for the groups of children included, and similar comparisons across groups were made in Studies II and IV. In Study III, intelligibility scores were also compared across three different listener conditions.

Three of the studies (Studies I, III and IV) aimed to assess validity, which can be done by examining whether the results from a test correlate with some other variable that is assumed to be related to the variable that the test is supposed to measure (Streiner & Norman, 2008). The related variable chosen here was the presence of a speech disorder, meaning that intelligibility scores for children with and without deviant speech were compared. More specifically, in Study I the intelligibility scores obtained for the children in the Control, Clinic and Clinic+ groups were compared while, in Studies II and IV, intelligibility scores for children in the SSD and TD groups were compared.

Validity was also investigated by means of an analysis of possible covariance between intelligibility scores and PCC scores, also assumed to be related to intelligibility. In Study IV, the validity of the STI-CH single-word test was additionally investigated by means of an analysis of the correlation with intelligibility in spontaneous speech. The ability of the two assessment methods (STI-CH and spontaneous speech, both transcribed orthographically) to correctly identify participants as regards group membership (SSD group or TD group) was analysed using discriminant function analysis (DFA), a statistical method which has previously been used in studies in the field of speech-language pathology to investigate the ability of a certain test to classify children with and without a speech or language disorder into the correct group (Bedore & Leonard, 1998).

Intra-listener reliability was analysed in Studies I, II and IV using different methods. Inter-listener reliability was analysed in all four studies; in Studies II–IV the tool used was intra-class correlation (ICC). The output of ICC is of two types: single measures and average measures. Single measures should be reported when an assessment method is intended for use by a single listener, e.g. in clinical work, whereas average measures are more appropriate when an assessment method is designed to be used in research, where the mean score of several listeners is frequently used (Shrout & Fleiss, 1979). Hence, single-measure ICC is indicative of the reliability of the scale when a sample is judged by a single listener, whereas average-measure ICC is indicative of the reliability of the scale when scores represent the average of different listeners' judgements. Since the assessment methods concerned are intended to be used both in clinical work (with one listener) and in research (with several listeners), both single and average measures were reported. Finally, the equivalence between pairs of lists in STI-CH was examined using correlation analysis and comparison of scores for the same child.



*Table 6. Overview of the statistical methods used to answer the research questions of the four studies included in the thesis*

RESEARCH OBJECTIVE	STUDY	STATISTICAL METHOD
Comparison of intelligibility scores obtained using the different assessment methods	I	Repeated-measures ANOVA Paired-samples <i>t</i> -test
Difference in intelligibility score between the different groups of participants	I	Kruskal-Wallis
	II,IV	Unpaired-samples <i>t</i> -test
	I,II & IV	Mann-Whitney U
Difference in intelligibility score between the three listening conditions	III	Repeated-measures ANOVA Paired-samples <i>t</i> -test
Prevalence of deviant speech in the 'good readers' and 'poor readers' groups	I	Fisher's exact test
Ability of the assessment of spontaneous speech/STI-CH to correctly identify participants as regards group membership	II, IV	Discriminant function analysis
Covariance of PCC and intelligibility (TD group)	II	Spearman's rank-correlation coefficient
Covariance of PCC and intelligibility (SSD group and whole group)	II	Pearson's correlation coefficient
Comparison of speech samples from the same child	II	Pearson's correlation coefficient
Covariance of STI-CH and PCC	IV	Pearson's and Spearman's correlation coefficients
Covariance of STI-CH and intelligibility in spontaneous speech	IV	Pearson's and Spearman's correlation coefficients
Intra-listener reliability	I	Point-by-point agreement
	II,IV	Pearson's correlation coefficient
	IV	Spearman's rank-correlation coefficient
Inter-listener reliability for the whole group	I	Pearson's correlation coefficient and $r^2$
Inter-listener reliability for the children with deviant speech	I	Spearman's rank-correlation coefficient
Inter-listener reliability	II, III & IV	Intra-class correlation
Equivalence of the two lists	IV	Pearson's and Spearman's correlation coefficients

## 5 RESULTS

The present chapter begins with a short introduction summarising the results. This is followed by four sections. In the first of these, data from Studies I and IV are presented. These data can be helpful when it comes to understanding the effect of speech material, elicitation technique and listener task on intelligibility scores and on the reliability and validity of a test. This relates to research question I. The next section presents data from Study II and describes aspects of the reliability and validity of the method for the assessment of intelligibility in spontaneous speech which was designed as part of the work on this thesis. This relates to research question II. The third section reports data from Study III in relation to how intelligibility scores based on spontaneous speech are affected by the number of times that the speech material is repeated to the listener. This relates to research question III. Finally, the last section presents results from Study IV that make it possible to describe the validity and reliability of the STI-CH single-word test, which was developed as part of the work on the present thesis. This last section relates to research question IV.

The pooled results from the four studies show statistically significant differences between various methods of assessment – multiple-choice assessment of single words, transcription of single words, transcription of sentences and transcription of spontaneous speech (Figure 1). Intelligibility scores were higher for spontaneous speech than for structured speech materials (single words and sentences) (Studies I and IV) and increased with the number of repetitions (Study III). Further, intelligibility scores were lower for the children with speech disorders than for the children with typical speech and language development, which supports the validity of the assessment methods used (Figure 1). The difference in intelligibility scores between good and poor readers found in Study I is potentially problematic in terms of validity, since it could suggest that low intelligibility scores may reflect poor reading skills rather than (or at least in addition to) poor speech functions. Reliability in terms of inter-listener reliability was satisfactory for all methods examined. However, structured speech materials appeared to yield higher reliability scores than spontaneous speech (Table 7).

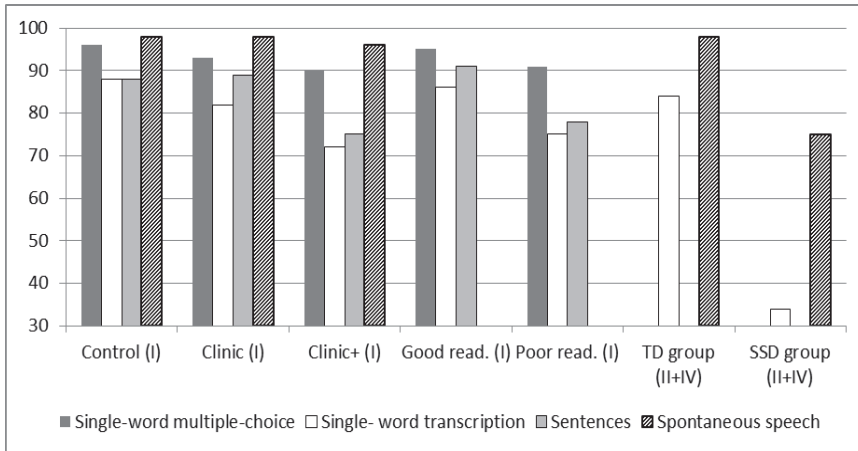


Figure 1. Average intelligibility scores obtained for the different assessment methods and speaker groups in Studies I, II and IV

Table 7. Overview of inter-listener reliability for the different assessment methods and speaker groups (study numbers in parentheses)

	All speakers (I)	Clinic+ (I)	All speakers (II & IV)	SSD group (II & IV)	TD group (II & IV)	Children with PCC < 90% (III)
Single-word multiple-choice		$\rho = 0.82$				
Single-word transcription	$r = 0.96$ $r^2 = 0.92$	$\rho = 0.95$	ICC = 0.97*	ICC = 0.93*	ICC = 0.90*	
Sentences	$r = 0.96$ $r^2 = 0.92$	$\rho = 0.93$				
Spontaneous speech	$r = 0.81$ $r^2 = 0.65$	$\rho = 0.80$	ICC = 0.71* ICC = 0.91**	ICC = 0.48* ICC = 0.79**	ICC = 0.48* ICC = 0.78**	ICC = 0.46–0.58* ICC = 0.80–0.93**

\*Single measure \*\*Average measure

## 5.1 Effect of speech material, elicitation technique and listener task

As regards speech materials, the highest intelligibility scores for the entire group of speakers in Study I was found for spontaneous speech ( $M = 97.4$ ,  $SD = 2.8$ , range: 88–100) and the lowest ones for single words by transcription ( $M = 81.5$ ,  $SD = 13.1$ , range: 42–98). The

corresponding comparison of results from Study IV for the transcription of spontaneous speech and single words yielded the same result: the average intelligibility score for spontaneous speech ( $M = 86.2$   $SD = 15.5$ , range: 51–100) was statistically significantly higher than that for single words ( $M = 58.9$   $SD = 30.6$ , range: 15–90):  $t(19) = 6.36$ ,  $p < 0.0001$ ;  $z = -3.920$ ,  $p < 0.0001$ . What is more, visual inspection of graphical representations of the results from Studies IV showed that each individual child had a higher intelligibility score for spontaneous speech than for single words (Figure 2).

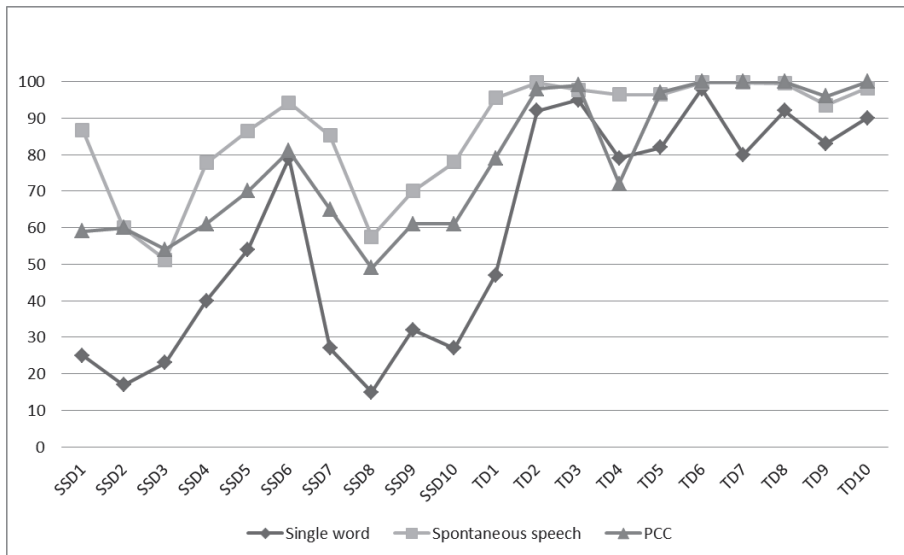


Figure 2. Scores obtained with the methods used to investigate validity in Study IV for individual children. The three lines represent intelligibility score based on single words, intelligibility score based on spontaneous speech (mean of the four listeners) and PCC score, respectively

The choice of elicitation technique had an impact on validity, since all methods based on reading (sentences and single words) had statistically significantly lower intelligibility scores for poor readers than for good readers. It thus appears that the assessment divided the children into groups based on their reading ability – a variable not assumed to be linked to speech intelligibility. However, one argument in support of the claim to validity can be made by reference to the fact that single words by transcription did yield a statistically significant difference between the group with manifest speech deviances according to SLP judgements and the group without them ('clinic+' vs. 'clinic'), but not between 'clinic' and 'controls' (where a difference was not expected since neither group had deviant speech).

When it comes to listener tasks, a mean difference of 12 percentage points was found between transcription and multiple-choice for single words, as regards the entire group of speakers in Study I. Words by multiple-choice did not give statistically significantly different results for the children with and without speech deviances, respectively, and nor did the intelligibility scores for spontaneous speech differ between those groups (probably owing to a ceiling effect).

## 5.2 Reliability and validity of assessment based on spontaneous speech

The inter-rater reliability of the method for assessing intelligibility in spontaneous speech investigated in Study II, which is based on ‘words perceived as understood’, yielded excellent ICC values for average measures and fair-to-poor ones for single measures (Table 7). Intra-listener reliability was high and significant ( $r = 0.94$ ). The majority (83%) of the cases (speaker–listener pairs) had a difference of less than 10 percentage points between the first and the second transcription. The intelligibility scores for the two different speech samples from the same child correlated statistically significantly ( $r = 0.73$ ) and 73% of the individual cases had a difference of less than 10 percentage points between the two samples. However, a few individual cases had a difference in excess of 20 percentage points.

As regards validity, there was a statistically significant difference between the group with speech-sound disorder ( $M = 74.7$ ,  $SD = 14.5$ , range: 51–94) and that without it ( $M = 97.7$ ,  $SD = 2.1$ , range: 94–100), and the discriminant function analysis yielded significant results with 17 of the 20 children classified as belonging to the correct group (three of the SSD children were classified in the TD group) (Figure 3). Further, intelligibility scores correlated statistically significantly with PCC scores for the overall group ( $r = 0.84$ ) and for each sub-group separately (TD group:  $\rho = 0.77$ ; SSD group:  $r = 0.79$ ).

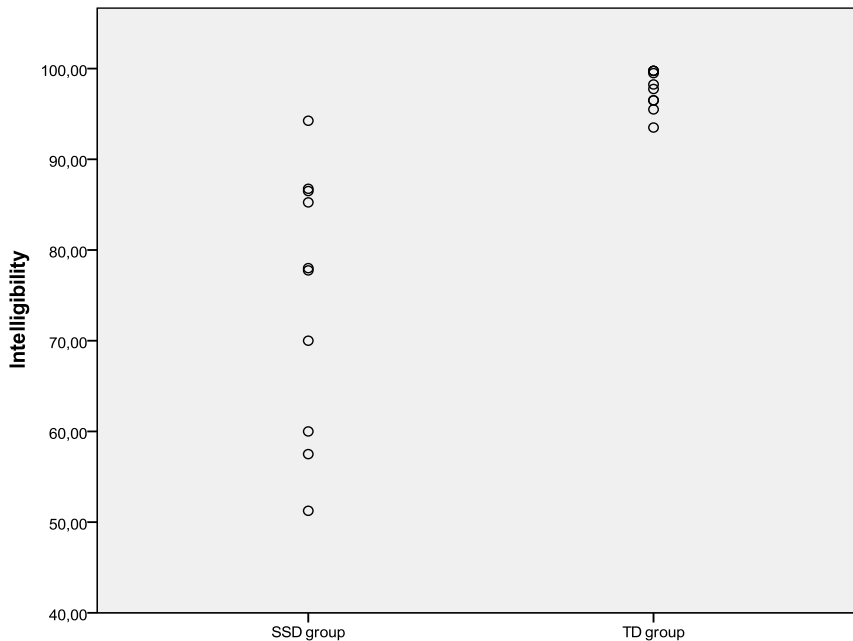


Figure 3. Intelligibility scores for the individual members of the two groups (SSD and TD) in study II.

### 5.3 Impact of the number of repetitions of the speech material

Statistically significant differences between intelligibility scores were found between the three conditions investigated in Study III: the intelligibility score increased with the number of repetitions. Inter-listener reliability in terms of ICC was approximately the same even though it was slightly higher for the third repetition. All three listener conditions had excellent inter-listener reliability (0.80–0.93) in terms of average-measure ICC and fair to poor inter-listener reliability (0.46–0.58) in terms of single-measure ICC. Visual exploration of the data revealed large differences in intelligibility scores across listeners; this variation between listeners was consistent throughout the three listening conditions for individual speaker–listener pairs. The children whose intelligibility scores differed the most across listeners all had PCC scores below the mean for the group (i.e. < 63). To illustrate this type of

listener variability, Figure 4 shows the intelligibility scores awarded to the two speakers S3 and S8 by the four listeners assigned to them.

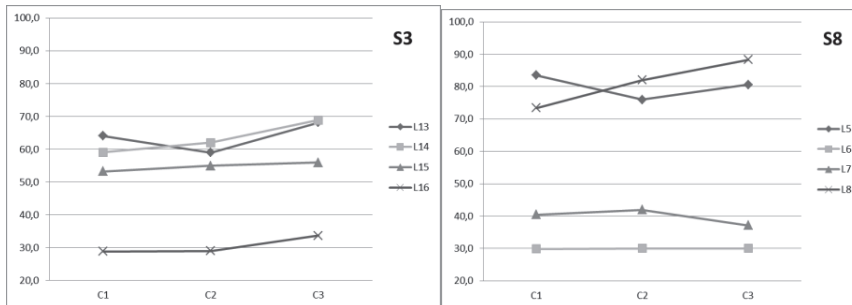


Figure 4. Listener variability for two of the speakers in Study III (S3 and S8). Intelligibility scores are plotted against the y-axis and listening conditions against the x-axis. Each line represents one listener

## 5.4 Reliability and validity of STI-CH

Each child completed two different word lists in STI-CH. Intra-listener reliability calculated separately for each of them was high ( $r = 0.94$  and  $r = 0.92$ ). Reliability was also high in terms of inter-listener reliability, with excellent single-measure ICC (0.97). Further, the two word lists yielded intelligibility scores that were equivalent for each child. The mean difference between two lists for the same child was 3.8 percentage points, and in 75% of cases the difference was 5 percentage points or less (Table 8).

As regards validity, a statistically significant difference was found between the SSD group ( $M = 33.9$ ) and the TD group ( $M = 83.8$ ). The discriminant function analysis yielded significant results, placing 18 of the 20 children in the correct group (one child in each group – SSD6 and TD1 – was misallocated). What is more, scores on STI-CH correlated statistically significantly with PCC scores ( $r = 0.94$ ) and with intelligibility scores on the spontaneous-speech task ( $r = 0.85$ ).

*Table 8. Equivalence of the different lists in STI-CH. Means of the intelligibility scores obtained by the four listeners for each child for each of the two lists and the difference between the scores for the two lists expressed in percentage points. Two of the word lists concerned are included as examples in Appendix 1; their ID numbers are shown in parentheses in the 'Percent' columns*

Speaker	First list		Second list		Difference
	Raw score	Percent	Raw score	Percent	Percentage points
SSD1	12	20	17	29 (848)	9
SSD2	9	15	12	20	5
SSD3	15	25	13	21	-4
SSD4	26	44	22	36	-8
SSD5	33	55	32	53	-2
SSD6	45	75	49	82	7
SSD7	17	28	16	26	-2
SSD8	10	16	8	13	-3
SSD9	19	32	19	32 (549)	0
SSD10	16	27	16	27	0
TD1	28	46	28	47	1
TD2	56	94	53	89	-5
TD3	56	93	58	96	3
TD4	44	73	51	85	12
TD5	51	85	47	78	-7
TD6	59	99	58	97	-2
TD7	49	82	47	78	-4
TD8	55	92	55	92	0
TD9	50	83	49	82	-1
TD10	55	91	53	88	3



## 6 DISCUSSION

The four studies included in this thesis explored different aspects of intelligibility and the assessment of intelligibility in children. As regards the influence of the assessment method and the elicitation technique, the results of Study I showed that the type of speech material used and the specific listener task performed played an important role for the resulting intelligibility score obtained by a child. It was clear that there was a need to develop methods specifically suited for children. Because of this, two methods were developed and further explored: one based on spontaneous speech and one on single words. These two assessment methods were explored in detail in Study II and Study IV, respectively.

The in-depth investigation in Study II of the assessment method based on the proportion of syllables perceived as understood in the transcription of spontaneous speech found this method to have adequate-to-high validity and reliability. It was suggested that this method is more time-efficient than ones where transcriptions are compared with a key script, meaning that it should be an attractive option for researchers and clinicians assessing intelligibility on the basis of spontaneous speech – as an alternative both to key-script methods and to the (less reliable) rating scales often used with spontaneous speech.

At this point, attention shifted to the listener and the listener task. In Study III, there was found to be a statistically significant yet small difference in intelligibility scores depending on how many times the listener heard the speech material. Reliability remained uniformly high regardless of the number of repetitions, suggesting that a single repetition – the least time-consuming option – might be adequate. The most striking finding from that study, however, was the considerable variation in intelligibility scores across listeners. The reasons for and ramifications of this will be discussed below.

Finally, pursuing the need that had become evident at the beginning of the work, the second assessment method developed and investigated (in Study IV) – the Swedish Test of Intelligibility for Children (STI-CH), a test material based on single words – yielded satisfactory results in terms of reliability and validity. All of these findings will be described in greater detail below, and there will also be a discussion of the concept of ‘intelligibility’ and the use of the term.

## 6.1 Effect of speech material and elicitation technique

In Study I, spontaneous speech yielded the highest intelligibility scores: the mean score was 16 percentage points higher than that obtained using transcription of single words. In Study IV, the scores for spontaneous speech were as much as 27 percentage points higher than those for single words. This is probably due to the additional cues provided by continuous speech, such as contextual and prosodic information, and it is in line with earlier findings (Gordon-Brannan & Hodson, 2000; Hodge & Gotzke, 2007). Separate analysis of the scores of the children with and without speech problems showed that the difference between single words and spontaneous speech appeared to be similar for the children with typical speech and language development in Study I (10 percentage points) and those in Study IV (14 percentage points). Such a difference may thus be expected, and it is indeed comparable to findings from earlier studies. For example, Hodge and Gotzke (2007) found a corresponding difference of 6 percentage points, and Gordon-Brannan and Hodson (2000) obtained 11 percentage points higher scores for spontaneous speech (using multiple-choice response as the listener task). While it is of course important to consider the differences between these studies in terms of the age of the participants and the methods used, it can reasonably be claimed that a difference of at least 10 percentage points or so can be expected between these two types of speech material, at least in children with typical speech and language development.

The earlier studies referred to also included children with various types of speech deviances, but it is not really possible to make comparisons in relation to them since the differences in the types and degrees of the speech problems involved are so large. Even so, one conclusion that can be drawn from the comparison of intelligibility scores from the different studies is that the difference between scores based on spontaneous speech and scores based on single words is larger for children with speech disorders than for children without them. In the two studies included in the present thesis, the difference between mean scores on single words and spontaneous speech, respectively, for the children with speech deviances was 24 percentage points (Study I) and 41 percentage points (Study IV). In fact, as shown in Figure 2, although every single child in Study IV had a higher score for spontaneous speech than for single words, the difference is smaller for the children with a less significant impairment to their intelligibility. The two earlier studies by Hodge & Gotzke and Gordon-Brannan & Hodson also displayed this trend, but not so clearly. To conclude, it is thus important to be wary of comparing

results obtained using different speech materials – especially as regards children with speech disorders, where the difference between speech materials seems to be particularly pronounced. Moreover, this underlines the importance, for study authors, of specifying in detail what type of material has been used.

There was also a difference between word lists and sentences in Study I, with higher scores for sentences than for single words (both by transcription). This is in line with other studies. The usual explanation is that sentences provide more contextual information (Osberger, 1992). However, it is important to note that the sentences in SWINT (Lillvik et al., 1999) are nonsensical sentences and thus do not contain much by way of contextual cues. This means that the reason for the higher scores obtained on sentences is not obvious; it could be, though, that the listeners did obtain some help from the grammatical context. The relationship between single words and sentences is less uniform as regards adult speakers with dysarthria. It is usually the case that severe dysarthria results in higher scores on words than on sentences whereas the opposite is true of people with mild dysarthria. In a previous study using the same assessment method (SWINT) on adults with dysarthria and a control group, sentences yielded lower intelligibility scores, and a wider range of scores, than word lists (Lillvik et al., 1999). Persson et al. (2003) reported 8 percentage points lower scores for sentences than for words in SWINT, for a group of speakers with 22q11 deletion syndrome aged 10–33, median age 14:1. This probably reflects difficulties in speech production rather than the amount of cues available to the listener and is in line with what is stated by Namasivayam et al. (2013) in relation to children with speech problems, namely that sentences represent greater linguistic and suprasegmental challenges to speakers than single words do, and that this additional load could affect articulation negatively.

This thesis used three different techniques to elicit speech for assessment of intelligibility: reading out loud (Study I), repeating after a model (Study IV) and speaking freely, i.e. spontaneous speech (Studies I–IV). Reading proved not to be suitable for children even as old as ten. In fact, elicitation by reading seemed to have low validity since it separated children who were good and poor readers (which it was not supposed to) as well as children with and without speech deviances (which it was). It is suspected that the reason why the validity of intelligibility scores based on reading is low is that they reflect the children's reading abilities rather than their ability to make themselves understood in speech-based communication. This question was raised in the previous study by Persson et al. (2003), where words and sentences from SWINT were also read out loud in a group including both schoolchildren and

adults with 22q11 deletion syndrome, some of whom also had cognitive delay. The overall conclusion must be that reading should be avoided in children and that repeating after a model is a more suitable option. In fact, the results from the method where repeating after a model was used as the elicitation technique (Study IV) showed that this method appeared to be valid both on the basis of a comparison between groups with and without speech disorder and on the basis of the relationship between the intelligibility scores obtained and other variables related to intelligibility in single words (i.e. PCC and intelligibility in spontaneous speech).

As already mentioned above, the method where the children spoke freely yielded considerably higher intelligibility scores than the structured speech materials in the present studies. Apart from the potential contributory factors mentioned above, it could also be the case that children consciously or unconsciously avoid words that they find difficult when speaking freely. It is not yet clear how this factor could be controlled for. Even so, the relatively high ecological validity of spontaneous speech compared with structured speech materials entails that it is worth including in an assessment. On the other hand, one important advantage of a structured speech material is that the level of complexity and the amount of speech material can be controlled for. Hence, a combination of a structured material and a spontaneous-speech material would appear to enable a comprehensive assessment of intelligibility in children.

## 6.2 Listener and listener task

The studies included in the present thesis used two different sets of listeners. Study I used two SLPs with experience in the field of cleft palate while Studies II–IV used a group of 20 SLP students. All listeners were female and between 20–45 years old, had normal hearing and Swedish as their first language. In other words, the listeners were rather similar, and indeed none of the studies was designed to study differences related to listener characteristics.

The two listeners in Study I were very similar in their assessments – i.e. inter-listener reliability was high, with point-by-point agreement between 88% and 95% for all assessment methods. However, in Study III it became evident that the transcriptions made by some of the listeners yielded intelligibility scores that differed markedly from those of other listeners. This finding is in line with previous research suggesting that listener variability is due to some variable other than sex, age or experience (McHenry, 2011; L. Pennington & Miller, 2007). The listeners were also very consistent in their performance –

those who obtained low intelligibility scores to begin with did not seem to benefit more than those who started out at a higher level from hearing an utterance again.

It can be noted in passing that this has a clinical implication: intelligibility does not seem to benefit from speakers repeating the same statement several times without expressing themselves in a different way or making some other change. Considering that reliability was comparable for all three conditions in Study III (hearing an utterance once, twice or three times), this also represents an argument in favour of the time-saving approach of using a single repetition in clinical contexts.

Given that the children with the lowest PCC scores in Study III gave rise to the greatest listener variation, it seems that in the absence of reliable acoustic-phonetic information, listeners differ in the strategies they use – or in the ability they possess – to perceive and make transcriptions of what is said. This type of ability has been discussed in relation to non-speech auditory, cognitive and working-memory capacities (Erb, Henry, Eisner, & Obleser, 2012; Ljung, Israelsson, & Hygge, 2013). The large variability of listeners' transcriptions found here underscores the importance of using the original observer at follow-up in clinical practice, unless it is possible to use the average score of perhaps two or preferably three listeners. For research purposes, it is important to be aware that individual results from different listeners can differ greatly, and thus to use averages for fairly large groups of listeners and/or to exclude listeners with extreme values or account for them separately. For a similar discussion, see Pennington & Miller 2007.

In Study I, two different listener tasks were combined with the speech material consisting of single words: multiple-choice and transcription. The multiple-choice task yielded an overall mean intelligibility score which was about 12 percentage points higher than that of the transcription task (the difference was 18 percentage points for children with speech deviances and 8 percentage points for the controls). Such a difference is expected, both because guesses are simply less likely to be wrong if the number of options is limited (i.e. an effect of chance-level performance) and because the listeners are guided in their perception by the different options given in the multiple-choice task. Other studies have found similar results. In an early study by Yorkston & Beukelman involving dysarthric speakers, the mean score on a multiple-choice task based on single words was 40 percentage points higher than the mean score on a transcription task (Yorkston & Beukelman, 1978).

Another aspect that must be taken into consideration when it comes to the listening task is whether the listeners are instructed to guess or not. In Study I, it was found that inter-listener reliability did not reach an acceptable level for spontaneous speech when guesses were included. This is in line with the manual of the Weiss intelligibility test, where listeners are not allowed to guess (Weiss, 1982), but other studies have encouraged guessing (Hodge & Gotzke, 2007; Tjaden & Liss, 1995). To this should be added that the concept of ‘guessing’ may actually be problematic in and of itself. When serving as test leader in Study III, the author of the present thesis noticed that the instruction not to guess elicited a great many questions from some (but not all) of the listeners, such as ‘How certain do I have to be not to call something a guess?’

The effect on intelligibility scores of the number of repetitions to the listener was investigated in Study III. A statistically significant but relatively small difference was found between the three conditions used (hearing each utterance once, twice or three times). This factor has not been investigated earlier, but it can be seen as an aspect of familiarisation, which has been studied previously (Borrie et al., 2011; Hustad & Cahill, 2003). Those earlier studies examined whether listeners obtain higher intelligibility scores for new material from a speaker whom they have already listened to. The typical aim of such studies has been to explore the possibility of training the communication partners of speakers with reduced intelligibility in order to enhance their intelligibility (Borrie et al., 2011) and to explore the effect of hearing the same speaker several times in the context of pre- and post-treatment studies (Hustad & Cahill, 2003). The present study differs from those studies in that it investigates scores for the same speech material. Knowing the effect of the number of repetitions is important to be able to decide how many times a listener should be allowed to listen to a speech material, and to determine whether the number of repetitions is of importance when results from different studies are compared. It turned out that although it was a statistically significant difference, the scores obtained were quite similar across the conditions. There was an increase of about 3 percentage points with each repetition, and hence an increase of 6 percentage points from the first to the third repetition. Hustad reported an increase of about 11 percentage points, with increases of 3–6 percentage points for each time and 9 percentage points between the first and the third time.

Borrie et al. (Stephanie A. Borrie et al., 2011) explored the impact of familiarisation with similar speech prior to intelligibility assessment in a study of dysarthric speech. In the familiarisation phase, one group of listeners heard dysarthric speakers (other than those whose speech they would later

transcribe) read a text ('passive familiarisation') and another group did the same while also having access to the text in writing ('explicit familiarisation'). A control group heard the same text read by neurologically intact speakers. In the subsequent transcription phase, a statistically significant difference in intelligibility scores awarded was found between the two conditions involving familiarisation with dysarthric speech and the control condition.

To sum up, both listener familiarisation in general and the number of repetitions have a significant impact on intelligibility scores. This means that the number of repetitions used should always be reported in studies including intelligibility assessment. Further, as was found in Study III, the number of repetitions does not appear to have any significant impact on the reliability of the intelligibility assessment or on the level of the intelligibility score, and so it does not matter greatly what number is chosen. From the perspective of time efficiency and listening burden, therefore, a single repetition would appear to be the best choice.

The results from Study III can also be discussed from the perspective of bottom-up and top-down processes in the brain. Since the listeners had heard all of the utterances produced by a given child when they started listening to those utterances for the third time, they knew the semantic context of the utterances (at least to the extent that they had been able to understand the general tenor of the child's speech). This could be expected to result in a larger increase in intelligibility scores between the second and the third repetition, especially for listeners whose initial scores were very low. However, it seems that once a listener has made a decision about an utterance on a semantic level, the bottom-up process may not be 'allowed back in' for a new attempt. As pointed out by Kent, 'speech perception is at times loosely derived from the acoustic signal. Knowledge-based hypotheses are pervasive influences' (1996, p. 8).

## 6.3 Reliability and validity

The issues of reliability and validity have been discussed to some extent in the previous sections, but they are further elaborated upon in this section. First of all, it is worth noting that in Study I the difference in intelligibility score between the first and second assessments by the same listener of the same speech sample from a child, when using assessment methods based on transcription, exceeded 10 percentage points. This should be taken into account as a possible confounder in clinical situations, where a clinician may evaluate material recorded before and after an intervention process.

Another important aspect of reliability is the difference between scores for different speech samples from the same child. This is crucial in cases where intelligibility based on single words or spontaneous speech is measured longitudinally or in relation to therapy, to determine whether a change in scores reflects a real change or is an artefact of the different samples. In Study II, it was found that the scores obtained using the spontaneous-speech sample from one child could differ by 0–24 percentage points ( $M = 9.4$ ,  $SD = 7.6$ ), while the differences between the two different word lists were in the range of 0–12 percentage points ( $M = 3.4$ ,  $SD = 3.3$ ). Possibly, the difference required as evidence that a real change has occurred should be larger for spontaneous speech than for word lists. A general conclusion to be drawn is that it may be preferable, both in clinical work and in research, to use both types of material to determine whether a change has actually taken place, and also to supplement them with a metric that measures a more functional type of intelligibility, such as the one developed by McLeod (McLeod et al., 2012).

Both intra-listener and inter-listener reliability was high for all speech materials and tasks in the studies included in the present thesis. However, there was less agreement between the two listeners for spontaneous speech when only the children with deviant speech were included. This is in line with earlier research (Whitehill & Chau, 2004), albeit using different speech material. When the method using spontaneous speech was further developed in Study II, using the syllable count instead of the word count to calculate the intelligibility score, inter-listener reliability in terms of average-measure ICC was found to be excellent (0.78–0.91) even though there were more listeners as well as greater variation in the speakers' speech deviances. In other words, this method is most certainly reliable for use in research where the key variable consists of a mean across several listeners (Shrout and Fleiss, 1979). However, in terms of single-measure ICC, which is more relevant when there is only one listener or judge, as is typically the case in clinical work, reliability was not quite so high: only 0.48–0.71. A previous study, where orthographic transcriptions of spontaneous speech were compared with a key script, reported single-measure ICC values ranging from 0.86 to 0.99 (Hodge & Gotzke, 2007). The STI-CH single-word test developed for Swedish children had sufficient inter-listener reliability (single-measure ICC = 0.90–0.97) compared with previously presented intelligibility tests based on single words, for which ICC values in the range of 0.90–0.99 (Hodge & Gotzke, 2007; Zajac et al., 2010) have been reported.

In terms of intra-listener reliability, the spontaneous-speech task explored in Study II had  $r = 0.94$ , which is satisfactory compared with Gordon Brannan



& Hodson (2000), who reported an intra-listener reliability for spontaneous speech in the range of  $r = 0.91-1.00$ . The intelligibility score was higher the second time round for the majority of the speech material. This could be due to an effect of familiarisation (Hustad & Cahill, 2003), yielding higher results when a listener hears an utterance for the second time. Unfortunately, comparisons with the values reported in the test manual of the Weiss intelligibility test (Weiss, 1982) are not feasible since it is not clear which of those values relate to the single-word task and the spontaneous-speech task, respectively. As regards the intra-listener reliability of the single-words STI-CH test,  $r$  was found to be 0.94 or 0.92, depending on the list, which is similar to values reported for previous tests, ranging from 0.74 to 0.98 (Gordon-Brannan & Hodson, 2000; Morris & Wilcox, 1999; Morris et al., 1995; Zajac et al., 2010).

In research as well as in clinical work, there is a need to ensure that the measurements and test instruments used really capture the intended variable, i.e. that they have high validity. In this thesis, the first investigation of validity related to the use on children of a word and sentence test designed for adults (SWINT) (Lillvik et al., 1999) and to an assessment method for spontaneous speech inspired by the one described by Weiss (1982). It turned out that SWINT did not only measure aspects of the speech disorder, which is expected and desirable in relation to intelligibility, but also measured aspects of reading skills. Hence it was concluded that SWINT did not have good validity for children and that a test designed specifically for children needed to be developed. The results from the other investigation also showed unsatisfactory validity, since the assessment method examined failed to distinguish between the groups with and without speech disorder, but it turned out that this could be explained by reference to the presence of a ceiling effect.

By contrast, the new single-word test subsequently developed (STI-CH) was found to have high validity when tested in Study IV since there was a strong correlation between the intelligibility scores obtained and other variables assumed to be related to intelligibility (PCC and intelligibility in spontaneous speech) and since the test proved to be able to distinguish between the groups with and without a speech disorder. The new method for spontaneous speech which was tested in Study II also turned out to have good validity in that intelligibility scores correlated strongly with PCC scores and that the scores of children with and without speech disorder differed clearly. To this can be added that the intelligibility scores obtained using that method and STI-CH were both in line with previous studies assessing intelligibility at word level and spontaneous-speech level (Flipsen, 2006; Gordon-Brannan & Hodson,

2000; Hodge & Gotzke, 2007; Weiss, 1982; Zajac et al., 2010), which constitutes another indicator of good validity.

## 6.4 Clinical and research implications

*'The idea that a single intelligibility score can be ascribed to a given individual apart from listener and listening situation is somewhat a fiction.'* (Kent, 1994, p. 81). While this is doubtless so, it is also the case – as Kent indeed concurs – that it is still possible and meaningful to assess intelligibility in children, provided, of course, that any differences in results for a child in clinical assessment are interpreted with caution and with consideration of the possible impact of different types of speech materials (or of different samples of the same type of speech material) and different listeners.

The high validity and reliability demonstrated for the methods explored in this thesis support this contention. The STI-CH single-word assessment can be used in clinical work with a single listener, provided that the same listener makes any follow-up assessments. The method for assessing intelligibility in spontaneous speech presented here has two major advantages. First, the approach of counting the number of syllables in both intelligible and unintelligible sequences when calculating the intelligibility score circumvents the problem of estimating the number of words in unintelligible strings that has been discussed in earlier studies (Flipsen, 2006; Shriberg, Austin, Lewis, McSweeney, & Wilson, 1997b). Second, it is a less complicated and therefore probably more time-efficient method because it does not entail the need to prepare a key script. This probably makes it more attractive to researchers and clinicians as an alternative to the less-reliable and less-valid rating scales which have often been used (Schiavetti, 1992; Whitehill, 2002).

STI-CH and the syllable-based method for spontaneous speech can also be recommended for use in research contexts, preferably with several listeners assessing the same speech material so that a mean score can be obtained, or so that extreme values produced by individual listeners can be detected and handled in an appropriate manner. It is also important to stress that the present thesis represents only the initial stage of the development and evaluation of these tests. There is an obvious need for additional research on the syllable-based method and on STI-CH alike.

## 6.5 The concept of ‘intelligibility’

The present thesis, like previous research, highlights the importance of using the concept of ‘intelligibility’ in a well thought-out way. This is because the operationalisation of the concept influences the choice of assessment method, and this in turn affects the outcome in terms of the intelligibility score obtained.

There are two main options as regards the definition of ‘intelligibility’. The first is to use this term narrowly in the sense of ‘the degree to which the acoustic signal [...] is understood by the listener’ (Yorkston et al., 1996) and use other terms to refer to situations where more context is included, as suggested by Yorkston’s use of the term ‘comprehensibility’. The second is to use it in a broader sense but specify what is included, for example by using composite terms such as ‘contextual intelligibility’ (Hustad, 2008) or ‘functional intelligibility’ (McLeod et al., 2012).

In other words, what we have here is a distinction between signal-dependent intelligibility (where the message is perceived from the acoustic speech signal only) and signal-independent intelligibility (where the message is also perceived from non-verbal sources such as semantic context or gestures). This is an important distinction which has recently been highlighted by Miller (2013). The author of the present thesis is of the opinion that it may in fact be preferable to take the distinction one step further and use completely different terms, reserving ‘intelligibility’ for that which is transmitted by the speech signal (meaning that the presentation of the material to the listener would have to be through the auditory channel only). It would also be advisable to specify the type of material (words, sentences or spontaneous speech) used for the assessment, referring for example to ‘spontaneous-speech intelligibility’. This would reduce the risk of comparisons being made between different studies as if they studied the same variables when the phenomena under study are in fact not quite (or not at all) the same. As a consequence, another term (not including ‘intelligibility’) would have to be used when visual or other information is also provided to the listener. While no specific proposal for such an alternative term will be put forward at this point, it is important to emphasise the importance of continuing the discussion and of remaining vigilant with regard to the concept of ‘intelligibility’ and how it is used.

The clinical implication of this is that when a child’s speech as such is being evaluated, intelligibility assessment is the appropriate choice, but when a child’s (and its communicative partners’) communicative abilities in a more

general sense are being investigated, this is not necessarily the case. It must be stressed, however, that these two concepts and the assessment methods associated with them should complement each other, not compete with each other (Miller, 2013).

## 6.6 Limitations and future research

The studies included in the present thesis have some limitations. First, the groups of speakers included are rather small. Second, only two different types of speech problems are covered. In addition, it would have been appropriate to screen the listeners involved with regard to their hearing.

During the execution of those studies, a few ideas for future research have emerged. One is that the clinical applicability of the two assessment methods has to be investigated further. It is also important for these methods to be tested on children with various types of speech disorders. Further, it would be interesting to explore the possibility of using the STI-CH word-based test to identify specific phonological and articulatory difficulties associated with intelligibility. Another potentially interesting topic of research concerns the medium of transmission (audiovisual vs. audio-only), above all its influence on intelligibility, in relation to speakers with various types and levels of speech disorders. Last but not least, it is desirable to maintain focus on the role of the listener and to conduct further studies of listeners' strategies and abilities (e.g. their phonological ability), which may affect their ability to understand the spoken signal.

# ACKNOWLEDGEMENT

I would like to express my sincerest gratitude to the following people:

Christina Persson, my supervisor.

Lena Hartelius & Jakob Åsberg, my co-supervisors.

Anette Lohmander, who made this possible from the start.

All parents and children, who took the time and made the effort to participate.

All students who participated as listeners.

Andrea Börjesson and Anna-Karin Ahlman, for helping me with data collection.

Christina Havstam and Sofia Antonsson, for performing the extremely time-consuming work of making the transcriptions for Study I.

Emilia Carlsson and Ann Nordberg, for performing the assessment of reading skills; Emilia also for being the best office mate ever.

Jonas Lindh, for creating the software for STI-CH and for the spontaneous-speech transcription.

Stefan Stenfelt, for providing the word bank which served as the basis for the creation of STI-CH.

All of my fellow doctoral students at the Division of Speech and Language Pathology at the Sahlgrenska Academy.

Anna-Karin Norrman, for always being so helpful and caring.

All of my colleagues at the Division of Speech and Language Pathology, for making it such a fun and inspiring place to work at.

Johan Segerbäck, language-checker, for understanding what I mean and helping me express it.

Katja Laakso, Malin Lillvik and Christina Havstam, my colleagues, for fruitful discussions and publication-celebration ceremonies including champagne.

My mother, Ann-Marie Lagerberg, for being so supportive and patient; my wonderful sister, Malin Lagerberg, and her family; my brothers – you don't know how much you all mean to me.

All of my wonderful caring and supportive friends – you know who you are.

## REFERENCES

- Allen, M. M. (2013). Intervention efficacy and intensity for children with speech sound disorder. *Journal of Speech, Language, and Hearing Research, 56*(3), 865-877.
- Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: a discriminant function analysis. *Journal of Speech, Language, and Hearing Research, 41*(5), 1185-1192.
- Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: a review of experimental studies. *Journal of Speech, Language, and Hearing Research, 55*(1), 290-305.
- Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2011). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes, 27*(7-8), 1039-1055.
- Brousseau-Lapre, F., & Rvachew, S. (2013). Cross-linguistic comparison of speech errors produced by English- and French-speaking preschool-age children with developmental phonological disorders. *International Journal of Speech-Language Pathology*. Epub ahead of print.
- Case, C., Forsberg, G., & Uppman, A. (2009). *Frequency of Words and Word forms: Analysis of Childrens Conversations with each other within the age groups 5-8 and 9-12 years*. (Master), Linköping University, Linköping, Sweden.
- Chen, C. L., Chen, H. C., Hong, W. H., Yang, F. P., Yang, L. Y., & Wu, C. Y. (2010). Oromotor variability in children with mild spastic cerebral palsy: a kinematic study of speech motor control. *Journal of Neuroengineering and Rehabilitation, 7*, 54.
- Coplan, J., & Gleason, J. R. (1988). Unclear speech: recognition and significance of unintelligible speech in preschool children. *Pediatrics, 82*(3 Pt 2), 447-452.
- Diagnostic and statistical manual of mental disorders : DSM-5*. (2013). Arlington, Va. : American Psychiatric Association, cop. .
- Dzioba, A., Skarakis-Doyle, E., Doyle, P. C., Campbell, W., & Dykstra, A. D. (2013). A comprehensive description of functioning and disability in children with velopharyngeal insufficiency. *Journal of Communication Disorders, 46*(4), 388-400.
- Erb, J., Henry, M. J., Eisner, F., & Obleser, J. (2012). Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia, 50*(9), 2154-2164.
- Flipsen, P. (2006). Measuring the intelligibility of conversational speech in children. *Clinical Linguistics & Phonetics, 20*(4), 303-312.
- Garcia, J. M., & Dagenais, P. A. (1998). Dysarthric sentence intelligibility: contribution of iconic gestures and message predictiveness. *Journal of Speech, Language, and Hearing Research, 41*(6), 1282-1293.

- Gordon-Brannan, M., & Hodson, B. W. (2000). Intelligibility/Severity Measurements of Prekindergarten Children's Speech. *American Journal of Speech-Language Pathology*, 9(2), 141-150.
- Havstam, C., Sandberg, A. D., & Lohmander, A. (2010). *Communication attitude compared to speech assessments in 10-year-old children with cleft (lip and) palate*. 13(2), 156-164
- Hodge, M., & Gotzke, C., L. (2007). Preliminary results of an intelligibility measure for English-speaking children with cleft palate. *Cleft Palate Craniofacial Journal*, 44(2), 163-174.
- Hustad, K. C. (2008). The relationship between listener comprehension and intelligibility scores for speakers with dysarthria. *Journal of Speech, Language, and Hearing Research*, 51(3), 562-573.
- Hustad, K. C., & Beukelman, D. R. (2001). Effects of linguistic cues and stimulus cohesion on intelligibility of severely dysarthric speech. *Journal of Speech, Language, and Hearing Research*, 44(3), 497-510.
- Hustad, K. C., & Cahill, M. A. (2003). Effects of presentation mode and repeated familiarization on intelligibility of dysarthric speech. *American Journal of Speech-Language Pathology*, 12(2), 198-208.
- Hustad, K. C., Schueler, B., Schultz, L., & Duhadway, C. (2012). Intelligibility of 4 year old children with and without cerebral palsy. *Journal of Speech, Language, and Hearing Research*.55,1177-1189
- Johannisson, T. B., Wennerfeldt, S., Havstam, C., Naeslund, M., Jacobson, K., & Lohmander, A. (2009). The Communication Attitude Test (CAT-S): normative values for 220 Swedish children. *International Journal of Language and Communication Disorders*, 44(6), 813-825.
- Kent, R. A. (1992). *Intelligibility in speech disorders : theory, measurement and management*. Amsterdam ;: Benjamin.
- Kent, R. D. (1996). Hearing and Believing: Some Limits to the Auditory-Perceptual Assessment of Speech and Voice Disorders. *Am J Speech Lang Pathol*, 5(3), 7-23.
- Kent, R. D., Miolo, G., & Bloedel, S. (1994). The Intelligibility of Children's Speech: A Review of Evaluation Procedures. *American Journal of Speech-Language Pathology*, 3(2), 81-95.
- Kent, R. D., Weismer, G., Kent, J. F., & Rosenbek, J. C. (1989). Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, 54(4), 482-499.
- Klinto, K., Svensson, H., Elander, A., & Lohmander, A. (2013). Speech and Phonology in Swedish-Speaking 3-Year-Olds with Unilateral Complete Cleft Lip and Palate Following Different Methods for Primary Palatal Surgery. *Cleft Palate-Craniofacial Journal*.Epub ahead of print.
- Kwiatkowski, J., & Shriberg, L. D. (1992). Intelligibility assessment in developmental phonological disorders: accuracy of caregiver gloss. *Journal of Speech and Hearing Research*, 35(5), 1095-1104.



- Lillvik, M., Allemark, E., Karlström, P., & Hartelius, L. (1999). Intelligibility of dysarthric speech in words and sentences: development of a computerised assessment procedure in Swedish. *Logopedics Phoniatrics Vocology*, 24(3), 107 - 119.
- Lindblom, B. (1990). On the communication process: Speaker-listener interaction and the development of speech\*. *Augmentative and Alternative Communication*, 6(4), 220-230.
- Ljung, R., Israelsson, K., & Hygge, S. (2013). Speech intelligibility and recall of spoken material heard at different signal-to-noise ratios and the role played by working memory capacity. *Applied Cognitive Psychology*, 27(2), 198-203.
- Lohmander, A., Borell, E., Henningsson, G., Havstam, C., Lundeborg, I., & Persson, C. (2005). *SVANTE : svenskt artikulations- och nasalitets-test*. Skivarp: Pedagogisk design.
- Lundeborg, I., & McAllister, A. (2007). Treatment with a combination of intra-oral sensory stimulation and electropalatography in a child with severe developmental dyspraxia. *Logoped Phoniatr Vocol*, 32(2), 71-79.
- Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes*, 27(7-8), 953-978.
- McHenry, M. (2011). An exploration of listener variability in intelligibility judgments. *Am J Speech Lang Pathol*, 20(2), 119-123.
- McLeod, S., Harrison, L. J., McAllister, L., & McCormack, J. (2013). Speech sound disorders in a community study of preschool children. *American Journal of Speech-Language Pathology*, 22(3), 503-522.
- McLeod, S., Harrison, L. J., & McCormack, J. (2012). Intelligibility in Context Scale: Validity and reliability of a subjective rating measure. *Journal of Speech, Language, and Hearing Research*.55,648-656
- McLeod, S., Verdon, S., Bowen, C., & International Expert Panel on Multilingual Children's, S. (2013). International aspirations for speech-language pathologists' practice with multilingual children with speech sound disorders: development of a position paper. *Journal of Communication Disorders*, 46(4), 375-387.
- Miller, N. (2013). Measuring up to speech intelligibility. *International Journal of Language and Communication Disorders*.Epub ahead of print.
- Morris, S. R., & Wilcox, K. A. (1999). *The Children's Speech Intelligibility Measure*. San Antonio: Psychological Corp.
- Morris, S. R., Wilcox, K. A., & Schooling, T. L. (1995). The Preschool Speech Intelligibility Measure. *American Journal of Speech-Language Pathology*, 4(4), 22-28.
- Namasivayam, A. K., Pukonen, M., Goshulak, D., Yu, V. Y., Kadis, D. S., Kroll, R., . . . De Nil, L. F. (2013). Relationship between speech motor control and speech intelligibility in children with speech sound disorders. *Journal of Communication Disorders*, 46(3), 264-280.

- Osberger, M. J. (1992). Speech intelligibility in the hearing impaired: Research and clinical implications.. In K. R. D. (Ed.), *Intelligibility in Speech Disorders* (pp. 233-264). Amsterdam: John Benjamins.
- Pennington, B. F., & Bishop, D. V. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, *60*, 283-306.
- Pennington, L., & Miller, N. (2007). Influence of listening conditions and listener characteristics on intelligibility of dysarthric speech. *Clin Linguist Phon*, *21*(5), 393-403.
- Persson, C., Lohmander, A., Jonsson, R., Oskarsdottir, S., & Soderpalm, E. (2003). A prospective cross-sectional study of speech in patients with the 22q11 deletion syndrome. *Journal of Communication Disorders*, *36*(1), 13-47.
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Atten Percept Psychophys*, *71*(6), 1207-1218.
- Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In R. D. Kent (Ed.), *Intelligibility in Speech Disorders*. Amsterdam/Philadelphia: John Benjamins publishing company.
- Shriberg, L. D., Austin, D., Lewis, B. A., McSweeney, J. L., & Wilson, D. L. (1997a). The percentage of consonants correct (PCC) metric: extensions and reliability data. *Journal of Speech, Language, and Hearing Research*, *40*(4), 708-722.
- Shriberg, L. D., Austin, D., Lewis, B. A., McSweeney, J. L., & Wilson, D. L. (1997b). The speech disorders classification system (SDCS): extensions and lifespan reference data. *Journal of Speech, Language, and Hearing Research*, *40*(4), 723-740.
- Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological disorders III: a procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders*, *47*(3), 256-270.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420-428.
- Sitler, R. W., Schiavetti, N., & Metz, D. E. (1983). Contextual effects in the measurement of hearing-impaired speakers' intelligibility. *Journal of Speech and Hearing Research*, *26*(1), 30-35.
- Streiner, D., L., & Norman, G. R. (2008). *Health measurement scales : a practical guide to their development and use*. Oxford Oxford University Press.
- Tjaden, K., & Liss, J. M. (1995). The role of listener familiarity in the perception of dysarthric speech. *Clinical Linguistics & Phonetics*, *9*(2), 139-154.
- Unicomb, R., Hewat, S., Spencer, E., & Harrison, E. (2013). Clinicians' management of young children with co-occurring stuttering and speech sound disorder. *International Journal of Speech-Language Pathology*, *15*(4), 441-452.

- Weismer, G., & Martin, R. E. (1992). Acoustic and perceptual approaches to the study of intelligibility. In R. D. Kent (Ed.), *Intelligibility in speech disorders*. Amsterdam/Philadelphia: John Benjamins publishing company.
- Weismer, G., Yunusova, Y., & Bunton, K. (2012). Measures to evaluate the effects of DBS on speech production. *Journal of Neurolinguistics*, 25(2), 74-94.
- Weiss, C. E. (1982). *Weiss intelligibility test*. Tigard: OR:CC Publications
- Whitehill, T. L. (2002). Assessing intelligibility in speakers with cleft palate: a critical review of the literature. *Cleft Palate-Craniofacial Journal*, 39(1), 50-58.
- Whitehill, T. L., & Chau, C. H. (2004). Single-word intelligibility in speakers with repaired cleft palate. *Clinical Linguistics & Phonetics*, 18(4-5), 341-355.
- Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders*, 11(6), 499-512.
- Yorkston, K. M., & Beukelman, D. R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. *Journal of Communication Disorders*, 11(6), 499-512.
- Yorkston, K. M., & Beukelman, D. R. (1981). *Assessment of intelligibility of dysarthric speech*. Tigard: CC Publications.
- Yorkston, K. M., Strand, E. A., & Kennedy, M. R. T. (1996). Comprehensibility of Dysarthric Speech: Implications for Assessment and Treatment Planning. *American Journal of Speech-Language Pathology*, 5(1), 55-66.
- Zajac, D., Plante, C., Lloyd, A., & Haley, K. (2010). Reliability and Validity of a Computer Mediated Single-Word Intelligibility Test: Preliminary Findings for Children with Repaired Cleft Lip and Palate. *Cleft Palate-Craniofacial Journal*.48(5),538-548



# APPENDIX

Swedish Intelligence Test for children - SWITCH	
List number:	848
1 telefon	31 hit
2 mjölk	32 kuckeliku
3 flygplan	33 kycklingen
4 åsnan	34 spindel
5 sparar	35 slä
6 dra	36 ord
7 träffar	37 agneta
8 bricka	38 båda
9 diaman	39 haft
10 fälla	40 av
11 andedräkt	41 funkar
12 racerbåt	42 jättedålig
13 inte	43 klassrummet
14 tjugofyra	44 lillebror
15 kunde	45 sak
16 många	46 flygande
17 du	47 gångband
18 skämt	48 loss
19 skojar	49 nina
20 måste	50 petra
21 verktyg	51 förstöra
22 lilla	52 vad
23 sex	53 hästar
24 ändå	54 banan
25 från	55 gjorde
26 rymdskepp	56 mjölkpaket
27 party	57 studsboll
28 stavas	58 hittar
29 sista	59 tusentals
30 pratar	60 kör

**Swedish Intelligence Test for children - SWITCH**

List number: 549

1 rör	31 marängen
2 siktet	32 annan
3 själva	33 förresten
4 du	34 fill
5 blommorna	35 undra
6 ihåg	36 popcorn
7 mot	37 långt
8 berättade	38 grodan
9 spökskepp	39 spela
10 vagnen	40 eller
11 död	41 lego
12 gå	42 oboy
13 knasigt	43 plötsligt
14 hall	44 banan
15 kulspruta	45 ljus
16 helikopter	46 veta
17 uggla	47 arg
18 alls	48 rast
19 popcornpaket	49 körde
20 söndagen	50 kände
21 kristians	51 trasig
22 tack	52 oskar
23 också	53 hus
24 köra	54 nästan
25 emellan	55 jättesöt
26 stinker	56 femton
27 äkte	57 söt
28 träffar	58 kanske
29 andra	59 grejer
30 fel	60 hör