Evaluation of nasal speech
A study of assessments by speech-language pathologists, untrained listeners and nasometry

Karin Brunnegård

Umeå 2008
Copyright © 2008 Karin Brunnegård

Dept of Clinical Sciences, Speech and Language Pathology
Umeå University
SE-901 87 Sweden

Photo on cover: Carina Kristiansson

Printed by Print & Media, Umeå, Sweden
# TABLE of CONTENTS

**ABSTRACT** -5-  
POPULÅRVETENSKAPLIG SAMMANFATTNING -7-  
LIST of ARTICLES -9-  
WORD LIST/ORDLISTA (English – Swedish) -10-  
PREFACE -11-  
INTRODUCTION -12-  
BACKGROUND -12-  

- Normal speech production  
- Resonance disorders and velopharyngeal impairment  
  - Resonance disorders in cleft lip and palate  
  - Resonance disorders in other groups  
- Auditory perceptual assessments  
  - Perceptual assessment of cleft lip and palate and VPI  
- Reliability of perceptual assessment of nasality  
- Validity of perceptual assessments of nasality  
  - Untrained listeners in speech assessments  
- Instrumental assessment  
  - The Nasometer™ - an acoustic instrument  

**AIMS** -21-  
**METHODS** -22-  

- Participants  
  - Speakers  
  - Listeners  
- Assessment materials and instrumentation  
  - Speech stimuli  
  - Audio-recordings of perceptual speech stimuli  
  - Assessment form for expert and non-expert SLPs  
  - Assessment form for untrained listeners  
- Experimental procedures  
  - Auditory perceptual assessment  
  - Measurement of nasalance  
- Data analysis  
  - Reliability of auditory perceptual ratings for expert SLPs  
  - Comparison between expert SLPs and untrained listeners  
  - Swedish norms for the Nasometer™  
  - Comparison between perceptual ratings and nasalance scores  

**RESULTS** -29-  

- Reliability of auditory perceptual ratings for expert SLPs  
  - Inter-rater reliability  
  - Intra-rater reliability  
- Comparison between expert SLPs and untrained listeners  
  - Inter-group correlations  
  - Qualitative comparison of ratings  
- Swedish norms for the Nasometer™  
- Comparison between perceptual ratings and nasalance scores  
  - Correlational analysis  
  - Nasalance threshold for presence of perceived hypernasality  

**DISCUSSION** -34-  

- Reliability of auditory perceptual ratings for expert SLPs  
  - Clinical implications  
- Comparison between expert SLPs and untrained listeners  
- Swedish norms for the Nasometer™  
- Comparison between perceptual ratings and nasalance scores  

**CONCLUSIONS AND FUTURE RESEARCH** -41-
ABSTRACT

Excessive nasal resonance in speech (hypernasality) is a disorder which may have negative communicative and social consequences for the speaker. Excessive nasal resonance is often associated with cleft lip and palate, velopharyngeal impairment, dysarthria or hearing impairment. Evaluation of hypernasality has proved to be a challenge in the clinic and in research. There are questions regarding the accuracy and reliability of auditory perceptual evaluations of nasal speech, and whether instrumental measures can be used to improve the reliability of clinical evaluation. There is also the question of whether clinical evaluation reflects the impact of hypernasality in a speaker’s everyday life.

The purpose of this thesis was to evaluate the extent of reliability problems connected with auditory perceptual assessment of nasality in speech, to explore whether they might interfere with treatment decisions or have an impact in the everyday life of patients, and whether they can be effectively diminished by the use of nasometry.

Speakers with cleft lip and palate or velopharyngeal impairment formed the basis of the clinical population used in this study. Speech samples from 52 of these speakers, along with samples from a reference population of 21 speakers who did not have cleft palate, velopharyngeal impairment or speech disorders were used in perceptual evaluation tasks. Fourteen speakers from the clinical population and 11 from the reference population also underwent nasometric evaluation. A further reference population of 220 children from three Swedish cities, whose ages were consistent with those used for clinical checks of children born with cleft palate were assessed with nasometry to establish normative data for the Nasometer™. Perceptual speech assessments were conducted on hyper- and hyponasality, as well as audible nasal air emission and/or nasal turbulence, using 5-point ordinal scales. Listeners were SLPs experienced in the evaluation of cleft palate speech, non-expert SLPs and untrained listeners. Listening assessments were performed from audio recorded speech samples assembled in random order. Nasometry measures were made on three speech passages each with specific phonetic content, using the Nasometer™, model II.

Perceptual evaluation Results showed that for hypernasality assessment, 15% of hypernasality assessments had disagreements between expert SLPs that were potentially important for clinical decisions, as did 6% of assessments for audible nasal air emission and/or nasal turbulence. For nasality problems, a comparison of expert and untrained listeners showed that they generally agreed on which speakers were hypernasal and on the ranking of nasal speakers. All speakers that had been rated with moderate to severe hypernasality by expert listeners were considered by the untrained listeners as having a serious enough speech disorder to call for intervention. However, in the case of audible nasal air emission and/or nasal turbulence the expert listeners were more prone to notice this feature than the untrained listeners.

Instrumental evaluation The development of normative values for the three Swedish passages for the Nasometer™ (comparable to normative values in other languages) has provided a basis for use of instrumental measures in Swedish clinics, oral sentences mixed sentences nasal sentences. The measures showed no significant differences due to city, gender or age within an age range of 4-10 years. When
nasometry measures were compared with perceptual evaluation of speech samples from the same speakers, all correlations were moderate to good for expert SLPs and non-expert SLPs. The difference between correlations was significantly higher for expert SLPs than for untrained listeners.

Reliability figures for perceptual assessments for expert SLP listeners indicated that there were some cases where lack of reliability could affect clinical decision making. However, in the main, judgements of nasality problems made by clinicians had everyday validity. They reflected the impressions of the everyday listener, especially in regard to the need for intervention. The study also indicates that now that Swedish norms are available, the Nasometer™ might be useful as a complement to auditory perceptual clinical speech assessments in Swedish cleft palate clinics in order to improve reliability of clinical assessment.
POPULÄRVETENSKAPLIG SAMMANFATTNING

Utredning av nasalt tal – en studie med bedömningar av logopeder, otränade lyssnare och nasometri

Bakgrund: Det är svårt att bedöma nasalt tal, dvs när röstklangen blir annorlunda på grund av ökad eller minskad resonans i näshål an, detta är känt från tidigare studier. Särskilt öppen nasalklang (hypernasalitet) är svår att bedöma. I denna avhandling har det framkommit att trots att lyssnarna är specialiserade logopeder och metodiken har anpassats efter de rekommendationer som förts fram på senare år så är överensstämmelsen och sambandet mellan olika lyssnares bedömningar och för samma lyssnare vid upprepad bedömning inte alltid tillfredsställande. Orsaker till att det är svårt att bedöma nasalt tal är att begreppet öppen nasalitet är svårdefinierat, att det ofta förekommer samtidigt som andra talavvikelser och att det kan blandas ihop med andra röstkvaliteter.

Nasalt tal kan orsakas av medfödda defekter i gommen t ex gomspalt, neurologisk sjukdom eller hörselfnedstätning. Nasalt tal pga defekt i gomen kan inte åtgärdas med träning utan behandlas vanligtvis kirurgiskt. Logopedisk träning används ofta som ett komplement eller alternativt vid andra orsaker till nasalt tal. God diagnostik och dokumentation före och efter beslut och åtgärd är därför mycket viktigt. Det är t ex också viktigt att titta på gomsens utseende och funktion vilket görs med röntgenfilmning och filmning med fiberoptik.

Metod: I de studier som ingår i avhandlingen har olika lyssnargruppers bedömning av nasalt tal jämförts, det har också undersöks om det finns en nytta med akustisk mätning av nasal klang med ett instrument som kallas Nasometer™. Hur logopeders bedömningar överensstämmer med lekmäns (otränade lyssnares) bedömningar av nasalt tal har också genomförts.

Resultat & Slutsatser: Resultaten av dessa studier visar att specialiserade logopeder och lekmän verkar vara överens om vilka lyssnare som har mest öppen nasalitet och att dessa behöver hjälp med sitt tal. Däremot har både lekmän eller logopeder som inte har erfarenhet av arbete med patienter med nasalt tal svårt att skilja på de två typer av nasalitet som finns, öppen och sluten nasalitet.

Mätning med Nasometer™ har tidigare rekommenderats som komplement till lyssnarbedömningar. En av artiklarna stödjer detta och det rekommenderas att mätresultat används för att bekräfta lyssnarbedömningen, framförallt i svårbedömda fall, och ge ett mått på nasalklangen. Mätning kan också användas av den logoped som är osäker på sin bedömning.

Betydelsen av vilket material man använder för bedömning med lyssning respektive Nasometer™ undersökes och det framkom att det inte behöver vara identiskt talmaterial vilket tidigare har framförts. Vilken typ av talmaterial som är lämpligast att använda vid mätningen med Nasometer™ klargjordes också.
Normaldata för akustisk mätning av röstklang, för 6-10-åringar, samlades in inom ramen för denna avhandling. Inga signifikanta skillnader mellan kön eller tre stora dialektområden framkom.

Utifrån de samlade resultaten rekommenderas att beslut om behandling ska grundas på bedömning av två logopeder eller upprepade bedömningar av en logoped och kompletterande mätningar med nasometri.
LIST of ARTICLES


IV. Brunnegård K, Lohmander A, van Doorn J. (Manuscript) Comparison of assessments by speech pathologists, untrained listeners and measurements by the Nasometer™.

Articles I, II and III are reprinted with permission from the publisher.
### WORD LIST/ ORDLISTA

<table>
<thead>
<tr>
<th>English</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>assessment</td>
<td>bedömning</td>
</tr>
<tr>
<td>cleft palate</td>
<td>gomspalt</td>
</tr>
<tr>
<td>cleft lip and palate</td>
<td>läpp- käk- gomspalt</td>
</tr>
<tr>
<td>evaluation</td>
<td>bedömning, utredning</td>
</tr>
<tr>
<td>fricative</td>
<td>frikativa (ex f, s)</td>
</tr>
<tr>
<td>hard palate</td>
<td>hårda gomen</td>
</tr>
<tr>
<td>hypernasality</td>
<td>hypernasalitet, öppen nasalitet</td>
</tr>
<tr>
<td>hyponasality</td>
<td>hyponasalitet, sluten nasalitet</td>
</tr>
<tr>
<td>fetus/fetal development</td>
<td>foster/fosterutveckling</td>
</tr>
<tr>
<td>malformation</td>
<td>missbildning</td>
</tr>
<tr>
<td>pharynx</td>
<td>svalg</td>
</tr>
<tr>
<td>plosive</td>
<td>klusil (ex p, t, k)</td>
</tr>
<tr>
<td>soft palate</td>
<td>mjuka gomen (gomseglet)</td>
</tr>
<tr>
<td>speech disorder</td>
<td>talstörning</td>
</tr>
<tr>
<td>speech and language pathologist</td>
<td>logoped</td>
</tr>
<tr>
<td>velopharyngeal impairment</td>
<td>velofarynxinsufficiens, nedsatt gomfunktion</td>
</tr>
<tr>
<td>velum</td>
<td>mjuka gomen (gomseglet)</td>
</tr>
</tbody>
</table>
PREFACE

Shortly after I started to work as a speech and language pathologist at Norrland University Hospital, Umeå, I became a part of the team working with cleft lip and palate. The studies included in this doctoral thesis have sprung out of my work with children with cleft palate and children with resonance disorders. As sometimes happens I had more questions after working for some years than I had when I started out. A position for a doctoral student was announced when Professor Jan van Doorn came to Umeå in 2002. I was fortunate enough to be accepted for that position and thus had the opportunity to pursue studies in my field of interest. Most of my work has been academic over the last five years but I remained in my clinical position part time since the interaction between clinical work and academic studies is fruitful. My hope is that through the research described in this thesis I will have contributed to expanding the knowledge regarding assessment of resonance disorders.

Karin Brunnegård
November 2008
INTRODUCTION

Excessive nasal resonance causes hypernasality, a resonance disorder that is common in certain groups of patients seen at speech-language pathology clinics in our hospitals. Hypernasality can have a negative influence on intelligibility and on the listener’s perception of an individual speaking with a hypernasal voice. Excessive nasal resonance is often associated with cleft lip and palate (CLP), dysarthria or hearing impairment but there are also nasal resonance disorders no such known origin. For individuals with nasality it is important that assessment methods are well developed so that treatment decisions are soundly based. Even though there is plenty of research in resonance disorders there are still unanswered questions to resolve regarding reliability of assessment methods, the relationship between assessments by listeners and instrumental measures, and questions of the relevance of the assessment to everyday situations.

BACKGROUND

Normal speech production

In order to understand the nature of a speech disorder it is in order to first describe normal speech production. The basis of voice production is controlled exhalation of air from the lungs which causes the vocal folds to vibrate and produce sound. The air stream is shaped into specific speech sounds by the articulators. The tongue and the lips are the most evident articulators but also the soft palate plays an important role. In typical speech production the soft palate is lifted for a large part of the speaking time to close the opening between the oral and the nasal cavities. In Swedish, only the sounds /m/, /n/ and /ŋ/ (the nasal consonants) are produced with the soft palate in the lowest position. The other speech sounds (all vowels and the remaining consonants) are oral speech sounds. A high intra-oral pressure is necessary to produce many of the speech sounds, which is achieved with lifted soft palate which forces the air stream exhaled from the lungs to take the route through the mouth only. When the soft palate (the velum) is lifted the walls of the pharynx move towards the soft palate, i.e. to achieve velopharyngeal closure. Velopharyngeal function is thus very important to speech. We speak at a rate of 12-14 speech sounds each second (Bradley, 1995) which means that many small but swift movements are combined into complex movement patterns every time an individual speaks. For each speech sound the articulators have to be changed according to place and manner of articulation for that particular sound. It is also known that the velum is lifted higher for consonants than for vowels, and lifted higher for high vowels than for low vowels (Bell-Berti, 1993). A speaker without velopharyngeal impairment will have a variation in the position of the velum and the nasal resonance for the same oral speech sound. This is due to the fact that sounds adjacent to nasal sounds have more nasal resonance than the same sound in a different position in a word; i.e. coarticulation (Warren, Dalston, & Mayo, 1993).

Resonance disorders and velopharyngeal impairment

If the movement of the velum and/or the pharyngeal walls is not sufficient for closure a velopharyngeal impairment (VPI) is evident which can result in hypernasal speech, audible nasal air emission/nasal turbulence, and weak pressure consonants. Other
terms for the same phenomenon are velopharyngeal insufficiency and velopharyngeal inadequacy but the term velopharyngeal impairment will be used in this text since this is in accordance with the terminology of the World Health Organization (WHO, 2002) and also suggested by Kuehn and Moller (2000) in their state-of-the-art article on speech and language in the cleft lip and palate population. Thus, there is oral resonance for most of the speech sounds and only nasal resonance for three speech sounds, namely: /m/, /n/ and /ŋ/. With velopharyngeal impairment the speaker will have difficulties with producing the difference between nasal and non-nasal speech sounds as needed. Common signs of velopharyngeal impairment are hypernasality and audible nasal air emission and/or nasal turbulence as well as weak pressure consonants.

**Hypernasality** is evident when the voice is characterized by excessive nasal resonance; this is mainly evident in vowels but also in voiced consonants. The reason is that the velopharyngeal port is not closed during production of oral speech sounds and the sound resonates in both the oral and nasal cavities. This may be due to a larger opening or timing difficulties in opening and closing the velopharyngeal port (Dotevall, Lohmander-Agerskov, Ejnell, & Bake, 2002). In acoustic terms there is among other things wider bandwidth of the formants which gives them lower intensity; there are also extra formants, called nasal formants (Lindblad, 1992). **Hyponasality** occurs when there is reduced nasal resonance, such as the sound of a blocked nose associated with a nasal congestion due to a common cold. Besides a cold, hyponasality may be due to an anatomical condition such as a deviation of the nasal septum but may also be due to timing difficulties (Watson, 2001). **Mixed nasality** is when both excessive nasal resonance and reduced nasal resonance occur at the same time.

**Audible nasal air emission and/or nasal turbulence** describes the phenomenon when the air stream through the nose becomes audible due to friction when the air passes through a narrow passage in the nasal, velar and/or pharyngeal area. This can occur either as a diffuse sound that is called audible nasal air emission, or a more distinctive friction sound which is called nasal turbulence. Nasal turbulence is often due to a friction in a more narrow passage. Other terms for nasal turbulence are nasal rustle, nasal snort or velopharyngeal friction sound but the term nasal turbulence will be used in this text. Audible nasal air emission and/or nasal turbulence often co-occurs with the resonance disorder hypernasality.

The term nasality is regularly used as a general term that includes both hypernasality, audible nasal emission and nasal turbulence, as in the title of this thesis.

The occurrence of **weak pressure consonants** is a common symptom of VPI. Among consonants there are a group of consonants that require high intraoral pressure, in Swedish the plosives and the fricatives. When the velopharyngeal port is not adequately closed these sounds may have reduced pressure and thus sound unclear. Weak pressure consonants often co-occur with hypernasality.

Treatment for speech disorders related to VPI generally involves a surgical intervention and is rarely manageable with speech therapy alone.

The major focus of this study is hypernasality and its assessment, particularly for speech associated with cleft lip and palate.
Resonance disorders in cleft lip and palate

CLP is a congenital malformation which is due to incomplete closure of the lip and/or the palate during early fetal development; a cleft is thus an opening in the lip tissue and/or palatal tissue and bone. A cleft of the lip and/or palate may affect feeding, speech, dental development, jaw development, hearing and appearance. The impact may thus be both functional and aesthetic.

The lip and palate are formed by parts that typically join within the first 12 weeks of fetal development. Some children have a cleft in the lip or the palate and others have a cleft in both the lip and the palate. The cleft in the palate may include only the soft palate or both the soft and the hard palate (these are both parts of the secondary palate). The cause of CLP is only partly known but both environmental and genetic factors are involved (Lees, 2001). There are a number of known environmental risk factors such as smoking, drugs, alcohol and pesticides. CLP most frequently occurs as an isolated malformation but may also be associated with other congenital defects or as a part of a syndrome (Lees, 2001). The incidence of cleft lip and palate was 0.7 in 1000 live births in a study from the Stockholm area, and the same incidence for isolated cleft palate (Hagberg, Larson, & Milerad, 1998). The same study indicated that about 20% of the whole study group with CLP had additional malformations.

The cleft is surgically treated early in a child’s life. This is true for Sweden and many other countries with developed health care systems (Watson, 2001). There is a great variation in surgical schemas for the surgery of clefts; some close the cleft palate in two steps: first the soft palate and than the hard palate, others close the soft and the hard palate at the same time (Shaw et al., 2001). In Sweden six regional cleft lip and palate centres or craniofacial centres provide care for all children with CLP in the country. These centres are located in Umeå, Uppsala/Örebro, Stockholm, Linköping, Göteborg and Malmö. The cleft palate teams are multi-disciplinary in order to provide high quality treatment for each individual with a cleft. In Sweden half of the centres provide closure of the hard and soft palate in one session (12-18 months of age) and the other half close the soft and the hard palate in two steps (6-9 months and 2-3 years respectively). Any remaining cleft in the primary palate (the alveolar process) is closed at a later point in time around 7-9 years of age.

The velopharyngeal impairment that is often associated with CLP makes this group a large group within the group of speakers with velopharyngeal impairment that are seen at the hospital speech clinics. The group with CLP is therefore a group which is usually the focus of studies in resonance disorders. Speech disorders related to CLP are mainly related to the cleft in the secondary palate whereas an isolated cleft lip or cleft in the lip and the primary palate/alveolar process does not cause speech disorders.

Incidence of speech disorders in the cleft palate population depends on cleft type, surgical methods, general development of the child etc. A European multi-centre study, the Eurocleft study (Grunwell et al., 2000), concluded that at the age of 11 to 14 years most speakers in a group with unilateral cleft lip and palate (n=131) had achieved acceptable and understandable speech. Among these five percent had severe hypernasality and just over 20% slight hypernasality. There were some articulation disorders but mostly fairly mild variants. A UK multi-centre study (Sell et al., 2001) in a group with UCLP (n=218) indicated that at age 12, 18% had mild-severe hypernasality and 17% had at least one serious articulation error. A study of speech in
5 year old children with isolated cleft palate (Persson, Elander, Lohmander-Agerskov, & Soderpalm, 2002) found that children with cleft of the soft palate only and no additional malformations or syndromes had satisfactory speech and in the group of children with cleft of the soft and hard palate 31% had mild to severe hypernasality and 10-15% had audible nasal emission/nasal turbulence.

There is ongoing discussion on surgical methods and timing of surgery in cleft lip and palate. Speech status is one principal outcome measure and it is therefore important that reliable assessment methods are applied in outcome studies.

**Resonance disorders in other groups**

Neurological disorders may effect the speech mechanism and result in a motor speech disorder, e.g. dysarthria. Speakers with dysarthria may exhibit either hyper- or hyponasal resonance. This has for example been reported for speakers with multiple sclerosis (Hartelius, Runmarker, & Andersen, 2000) and after a cerebrovascular accident (Thompson & Murdoch, 1995). Anatomical restrictions after surgical treatment for cancer may also cause VPI and thus hypernasality in speakers (Borggreven et al., 2005). Speakers with hearing impairment are another group that may exhibit hypernasal speech which is most probably due to a lack of auditory feedback (Nguyen, Allegro, Low, Papsin, & Campisi, 2008).

Other reasons for an impaired velopharyngeal function are a congenital short velum or a congenital dysfunction in the movements of the velum and the pharyngeal walls. Furthermore, at times large pharyngeal tonsils hinder the movement of the velum which gives a secondary velopharyngeal impairment (Henningsson & Isberg, 1988).

**Auditory perceptual assessments**

The starting point for evaluation by a speech-language pathologist is usually an auditory perceptual assessment (Hartelius & Lohmander, 2008), and this is also the most common assessment in clinical settings (Kent, 1996). Thus, the ear and the ability to process and interpret what we hear is the most important assessment instrument (Moll, 1964). In an article by Kent (1996, p 7) it is stated that: “The ear is the essential tool of the speech-language pathologist. Auditory perceptual judgments are typically the final arbiter in clinical decision-making and often provide the standards against which instrumental (so-called “objective”) measures are evaluated.” Methods for auditory perceptual assessment in speech and language disorders include phonetic transcription, the use of rating scales to quantify speech and language features (such as hypernasality, audible nasal air emission and/or nasal turbulence, and intelligibility), and qualitative descriptions. In the assessment of nasality rating scales are usually used to quantify hyper- and hyponasality.

**Perceptual assessment of cleft lip and palate and VPI**

Assessment through listening is also the standard method for speech assessment in patients with cleft lip and palate (Folkins & Moon, 1990; Sell & Grunwell, 2001). The method for assessment at the cleft palate clinics in Sweden has been developed over the years through discussions and collaborations between the team speech and language pathologists. In 2005 a test material, Swedish test of articulation and nasality – SVANTE (Lohmander et al., 2005), with a comprehensive manual was released and is now used at the cleft palate clinics for evaluation of speech at all ages. The assessment of resonance and articulation is performed in one session which is
usually audio recorded to ensure the possibility of making a detailed transcription and allowing for later listening for clinical and research purposes. The test material includes single words, sentences and elicitation of continuous speech. This has been recommended in order to ensure a comprehensive evaluation (Kuehn & Moller, 2000; Sell, 2005). The nasality variables are rated on a 5-point ordinal scale. The scale is not an EAI-scale since there is no assumption on equal distances between the scale points. On the contrary, scale value 1 is described as a very slight deviation close to normal and also found in the normal population. There are thus three scale values at the low end of the scale, one in the middle and one at the high end of the scale. SVANTE has great similarities with the United Kingdom tests for assessment of cleft palate speech: Great Ormond Street Speech Assessment, GOS.SP.ASS.98 (Sell, Harding, & Grunwell, 1999) and Cleft Audit Protocol for Speech, CAPS (Harding, Harland, & Razzel, 1997). The methods are similar regarding the use of transcription, 4-5-point scale for assessment of nasality and audible nasal air emission and/or nasal turbulence and a comprehensive speech stimulus (words, sentences, spontaneous speech).

**Reliability of perceptual assessment of nasality**

The issue of reliability of ratings is very important in clinical and research assessments especially since hypernasality is a variable that has been shown to be difficult to assess reliably (Counihan & Cullinan, 1970; McWilliams, Morris, & Shelton, 1990; Keuning, Wieneke, & Dejonckere, 1999; Persson, Lohmander, & Elander, 2006). One reason for this is the influence of other co-existing speech variables on the perception of nasality. Variables that have been reported to influence the perception of, or are interrelated to nasality are: audible nasal air emission/nasal turbulence, articulatory proficiency, pitch and loudness (Fletcher, 1973; McWilliams et al., 1990; Zraick et al., 2000).

Two types of rater reliability are of interest: intra-rater reliability that indicates if a rater makes the same rating when rating the same speech sample more than once, and inter-rater reliability which indicates if different listeners give the same rating to a given speech sample. In contrast to the problems reported there are also a number of examples of good inter- and intra-rater reliability for hypernasality (Grunwell et al., 2000; Hayden & Klimacka, 2000; Pulkkinen, Haapanen, Paaso, Laitinen, & Ranta, 2001; Sell et al., 2001) but the reasons for the differences between studies remain unresolved.

In a comprehensive review of the limitations of auditory perceptual analysis Kent, 1996 informs us of several general problems with auditory perceptual assessments. These issues are relevant to the assessment of speech disorders related to VPI:

1) **Judges do not appear to have equivalent definitions of dimensions to be rated.**

The definitions of hypernasality, different degrees of hypernasality, audible nasal air emission/nasal turbulence etc are not very exact (Sweeney & Sell, 2008). Sometimes the phenomenon of audible nasal air emission and/or nasal turbulence is assessed under the variable of hypernasality (Paal, Reulbach, Strobel-Schwarthoff, Nkenke, & Schuster, 2005), in other studies hypo- and hypernasality are rated on the same scale representing opposite ends of a continuum (Hayden & Klimacka, 2000). Some authors include active nasal fricatives in the concept of audible nasal air emission (Kummer, 2008) and some use the concept nasal emission and include both audible and inaudible nasal emission ( Watterson, Lewis, & Deutsch, 1998; Kummer, 2008).

2) **Specialists fail to reach consensus on which perceptual dimensions should be**
rated for a given disorder  Professionals in the field, nationwide and internationally, need to work towards a consensus regarding what perceptual dimensions need to be rated and how these are defined (John, Sell, Sweeney, Harding-Bell, & Williams, 2006). Today there is some consensus regarding assessment protocols of speech associated with VPI. Most studies include assessment of hypernasality, audible nasal air emission and/or nasal turbulence and pressure consonants (Lohmander & Olsson, 2004; Henningsson, 2008). 3) Perceptual ratings of various dimensions are intercorrelated, that is they are not independent. It is true for speech disorders related to VPI that hypernasality, audible nasal air emission/nasal turbulence and weak pressure consonants are interrelated since they have the same origin and often co-occur (McWilliams, 1990). There is also a connection between ratings of hypernasality and pitch and volume (Zraick et al., 2000) 4) Differences among expert judges are larger than the differences needed for diagnostic classification or differences needed for evaluation of the effects of intervention. According to research (Kreiman, Gerratt, Precoda, & Berke, 1992) expert raters have differing internal standards for voice variables. This is important in outcome research and also in assessment of outcome for clinical purposes, i.e. when we assess speech pre- and post-treatment we might get false positive results when it is really due to a change of assessing clinician and if we compare results reported with different listeners we do not know if it is the listeners’ standards or the actual outcome that is different.

In a review of published studies one will find that scales with between 2 and 7 scale points are used in the assessment of hypernasality (Hardin, van Demark, Morris, & Payne, 1992; Pulkkinen, Haapanen, Laitinen, Paaso, & Ranta, 2001). Some argue that a scale with less scale points would increase inter- and intra-rater reliability (McWilliams et al., 1990; Pulkkinen, Haapanen, Laitinen et al., 2001). This can also be inferred from a literature review since two of the studies with the best rater reliability have included a scale with few scale points: a binary scale (Pulkkinen, Haapanen, Laitinen et al., 2001) and a scale with three scale points (Grunwell et al., 2000). Composite scores of VPI also seem to increase reliability (Park et al., 2000). Extensive training of research SLPs was included in another study with good rater reliability (Sell et al., 2001). The study by Grunwell et al. included blinded, randomised recordings and assessment by external raters whereas the studies by Park et al. and Pulkkinen et al. did not use external raters. In the study by Sell et al. the raters for the study made the recordings together with the SLP on the CLP-team but they were not involved in the treatment of the children.

In line with the identified problems suggestions have been made on how to improve rater reliability. Some are basic requirements such as good listening conditions and good quality of recordings. Other recommendations have been to use raters that are experienced in the field of resonance disorders (Hayden & Klimacka, 2000; Lewis, Watterson, & Houghton, 2003), to use listener training (John et al., 2006), anchor stimuli (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993), a correct type of scale (Whitehill, 2002) and to use more detailed definition of variables (Henningsson et al., 2008; John et al., 2006).

In many articles on cleft palate speech no report of intra- or interreliability is made or only a measure for either inter- or intra-rater reliability is reported (Lohmander & Olsson, 2004; Sell, 2005). The first step is to always report these measures.
Validity of perceptual assessment of nasality

The validity of the variable hypernasality may come into question when the reliability of ratings is not as high as one would wish. Fletcher (1976) performed a series of experiments that demonstrated that the entity of (hyper-) nasality is noticed by untrained listeners. Their experiments also showed that the listeners distinguished between levels of hypernasality.

Untrained listeners in speech assessments

Research within the field of speech pathology has added the use of untrained listeners to add everyday significance to the results. Studies on dysarthria (Dagenais, Watts, Turnage, & Kennedy, 1999; Dagenais, Brown, & Moore, 2006) suggest that ratings by professionals “do not necessarily reflect the attitudes of the community with whom the impaired speakers normally associate.” (p 142, 2006). The life of an individual with a speech disorder will be influenced by how people they meet in day-to-day situations perceive them and understand what they say.

The International Classification of Functioning, Disability and Health (ICF) published by the World Health organisation (WHO) is a tool to describe functioning and health in a comprehensive and multidimensional way. The latest version of the classification was published in 2002 (WHO, 2002). Functioning and disability are described as outcomes of interactions between a disorder and contextual factors (WHO, 2002). A disability is defined as a dysfunction at one or more of the levels of impairment, activity limitations and participation restrictions. Impairment would be a structural deficit such as a velopharyngeal impairment, activity limitations would be the speech disorder, participation restrictions would be limitation of participation in daily life due to a speech disorder causing reduced intelligibility. Social attitudes is an example of a contextual factor that also may influence participation. The authors of ICF emphasise the importance of contextual factors to functioning and disability, including both climate and terrain as well as social attitudes and laws. The ICF is useful as a framework for discussing speech disorders and the relevance of clinical assessments to the day-to-day life of patients.

The assessment of resonance disorders by a professional speech clinician will more likely assess the level of impairment or activity limitation, while the judgment by an untrained listener will presumably inform us if this speaker experiences restrictions in participation in various situations. Assessment by untrained listeners is used to investigate how clinical findings of hypernasal speech can be generalised to everyday life for the speakers: do the untrained listeners hear what the SLPs hear? If we find out how an untrained listener perceives the speech of patients with nasal speech this will increase our understanding of speech disability resulting from participation restrictions.

One could also describe the issue in terms of need of change. Is hypernasality something that needs to be changed? Is reducing hypernasality an appropriate goal for treatment? To find the level of hypernasality which requires intervention we may need to compare the perceptual assessments of the speech clinician with the perceptual judgements from untrained listeners to validate the SLPs’ judgements. There have been suggestions that SLPs perceive more abnormal variables than untrained listeners which might lead to overtreatment of some patients (Witt, Berry, Marsh, Grames, & Pilgram, 1996). The opposite has also been proposed i.e that
inexperienced listeners have a tendency to exaggerate the severity of the hypernasality (Lewis et al., 2003) or that the SLP underestimates the speech difficulties (Bagnall & David, 1988). Two studies have reported similarities between judgements by SLPs and untrained listeners (Starr, Moller, Dawson, Graham, & Skaar, 1984; Tonz et al., 2002). This calls for further investigations since it is the degree to which speech sounds deviant to an untrained listener that determines whether an individual’s speech is a problem (Shuster, 1993).

**Instrumental assessment**

The difficulties in auditory perceptual assessment have lead to a search for instrumental methods that can give reliable measures. In an state of the art article of cleft palate speech research (Kuehn & Moller, 2000) it was concluded that no instrumental technique can replace perceptual analysis but there are a number of choices for the clinician who needs to complement auditory perceptual assessment with instrumental assessments. There are variants of visual assessment that are very important in providing information about the status of the velopharyngeal function in relation to speech but which only indirectly measures speech. There are also aerodynamic measurements available but these will not be discussed in this text because the interest for this study was acoustics and auditory perceptual properties of speech. The focus of the current study is a non-invasive acoustic instrument, the Nasometer™ which was specifically developed in an effort to provide a reliable measure that is directly comparable with perceptual speech assessment.

**The Nasometer™ – an acoustic instrument**

The Nasometer™, the Nasality Visualization System (a development of the OroNasal System) and the NasalView are three acoustic instruments that measure nasalance, i.e. the proportion of nasal energy to the total acoustic energy in a speech signal. They are all computer based and measure acoustic energy in a similar manner (Kummer, 2008). The most widely used acoustic tool for assessment of hyper- and hyponasality is the Nasometer™ which is also the instrument that has most data published regarding its reliability, norms and comparability to perceptual assessment (Brunnegård & van Doorn, In press).

The Nasometer™ use two microphones separated by a plate held by a head set to record the nasal and the oral speech signal simultaneously. The recorded signal is analog but is converted to digital before the computation of the nasalance score is made. The signal is filtered to a 300Hz bandwidth signal with the centre around 300 Hz. The formula for calculation of nasalance is: (nasal energy/(nasal energy+oral energy)) x100. The Nasometer™ was a development of the Tonar (I and II) (S G Fletcher, Adams, & McCutcheon, 1989) and the current model is the second version, model 6400. Most published studies up to date have been using the first model, 6200, but more recent publications use Nasometer II (Watterson, Lewis, & Brancamp, 2005; Bae, Kuehn, & Ha, 2007; Lewis, Watterson, & Blanton, 2008). The clinician can use the nasalance score to compare with their own perceptual rating and to compare pre- and post treatment scores.

To make the nasalance scores useful in clinical practice there need to be normative data collected for the language in question. Normative data have been published for several variants of English (Dalston, Neiman, & Gonzalez-Landa, 1993; van Doorn & Purcell, 1998; Sweeney, Sell, & O’Regan, 2004) and for several other languages e.g.
Dutch, French, Japanese, Finnish (Tachimura, Mori, Hirata, & Wada, 2000; van Lierde, Wuyts, De Bodt, & Van Cauwenberge, 2003). There are no published data for any Scandinavian languages. There have been studies showing differences in nasalance values for language and dialect (Seaver, Dalston, Leeper, & Adams, 1991; Leeper, Rochet, & Mackay, 1992; Nichols, 1999; van Lierde, Van Borsel, Moerman, & van Cauwenberge, 2002) or age (Haapanen, 1991; Van Lierde et al., 2003; Hirschberg et al., 2006) and gender (van Lierde, Wuyts, De Bodt, & Van Cauwenberge, 2001; Prathanee, Thanaviratananich, Pongjunyakul, & Rengpatanakij, 2003;) but there are also findings that there are no differences due to dialect, age or gender (Kavanagh, Fee, Kalinowski, Doyle, & Leeper, 1994; van Doorn & Purcell, 1998; Nichols, 1999; Sweeney et al., 2004; Mishima, Sugii, Yamada, Imura, & Sugahara, 2007). Those that have found age differences has mainly compared children with adults and not groups of children or adults.

Phonetic content speech stimulus has heavy influence on nasalance scores, especially the inclusion of nasal phonemes (m, n, ŋ) (Watterson, Hinton, & McFarlane, 1996) and influence of high vowels (Lewis, Watterson, & Quint, 2000). The influence of vocal loudness has been investigated but not proven to have any impact (Watterson, York, & McFarlane, 1994). There is also intra-speaker variability between recordings that is important to be aware of. Within-person variability is considered within normal variation up to five points (Watterson et al., 2005) and for speakers with hypernasality variation is probably even greater. A combination of auditory perceptual and instrumental measures are logical to ensure good quality evaluations. Comparisons between auditory perceptual ratings and nasalance scores have been made and some studies have found good correlation (Dalston, Warren, & Dalston, 1991; Watterson et al., 1996; Hirschberg et al., 2006; Sweeney & Sell, 2008;) and others moderate (Dalston et al., 1993; Watterson, McFarlane, & Wright, 1993; Keuning, Wieneke, van Wijngaarden, & Dejonckere, 2002) even low correlation (Nellis, Neiman, & Lehman, 1992; Lewis et al., 2003).

In summary, it is well documented that assessment of hypernasality in speech is potentially unreliable, which has led to problems with comparison and interpretation of research. This is of particular importance for evaluation of results from studies that compare speech outcomes of surgical methods for cleft palate repair. The poor reliability also raises questions regarding clinical assessments that are the basis for decisions about speech treatment and surgery. It was the uncertainty associated with auditory perceptual assessment of hypernasality that led to the development of devices such as the Nasometer™ that directly and objectively measure the acoustic speech signal. However, it has been found that nasometry measures also have a degree of variability in terms of test-retest scores, which raises the question about whether the device can effectively be used to improve assessment reliability. It is of great importance to evaluate the extent of reliability problems connected with auditory perceptual assessment, to explore whether they might interfere with treatment decisions or have an impact in the everyday life of patients, and whether they can be effectively diminished by the use of nasometry.
AIMS

The overall aim was to investigate aspects of evaluation of nasality which was formulated into four aims:

- To investigate the reliability of expert SLPs’ auditory perceptual assessment of hypernasality and related speech characteristics.
- To compare the ratings by untrained listeners with ratings by expert SLPs for cleft palate speech.
- To validate nasalance scores from the Nasometer™ with perceptual assessments of hypernasality by expert SLPs, non-expert SLPs and untrained listeners.

In order to meet the third aim it was necessary to conduct a norms study for Swedish on the Nasometer™:

- To establish normative nasalance values as measured with the Nasometer™ II for Swedish speaking children and investigate if there were significant differences due to age, gender and regional dialect.
METHODS

Participants

Speakers

All four studies involved perceptual and/or instrumental evaluation of nasal qualities of speech. There were two main categories of speakers who participated: children with cleft lip and/or palate or VPI of other etiology and children with no known speech disorders. In those studies that required perceptual evaluation of speech (studies I, II, IV) there were a total of 52 children with cleft lip and palate or VPI and 21 children without speech disorder or cleft lip and palate. In the normative nasalance study (study III) 220 pre-school and school children without speech disorders participated. Table 1 summarises the characteristics of participants in the four studies.

Parents for all included speakers gave their consent after receiving written or written and oral information about the studies.

Table 1. Characteristics of speakers

<table>
<thead>
<tr>
<th>Speaker group description</th>
<th>n (F,M))</th>
<th>Age at time of recording</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unilateral cleft lip &amp; palate (UCLP)</td>
<td>12 (2F, 10M)</td>
<td>9-11</td>
<td>I</td>
</tr>
<tr>
<td>No language disorder or cognitive disabilities, no hearing impairment requiring hearing aid.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleft palate only (CPO)</td>
<td>26 (17F, 9M)</td>
<td>9-11</td>
<td>I and II</td>
</tr>
<tr>
<td>No language disorder or cognitive disabilities, no hearing impairment requiring hearing aid.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison group 1</td>
<td>10 (5F, 5M)</td>
<td>9-11</td>
<td>I and II</td>
</tr>
<tr>
<td>Age typical speech and no cleft palate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Velopharyngeal impairment due to CLP or other etiology</td>
<td>14 (10F, 15M)</td>
<td>6-16</td>
<td>IV</td>
</tr>
<tr>
<td>Symptoms of VPI with no moderate or severe articulation disorder.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comparison group 2</td>
<td>11 (5F, 6M)</td>
<td>7-17</td>
<td>IV</td>
</tr>
<tr>
<td>Age typical speech and no VPI or speech disorder.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normative study group</td>
<td>220 (128F, 92M)</td>
<td>4:0 to 5:11 (n=45)</td>
<td>III</td>
</tr>
<tr>
<td>Age typical speech and no speech disorder.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recruited in Göteborg, Stockholm, and Umeå.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*There were no known syndromes or additional malformations in the groups. Five children in study I and II with cleft of the hard and soft palate had Pierre Robin sequence.
Listeners

Both SLPs and untrained listeners participated in the perceptual studies. In this text SLPs working on a cleft palate or craniofacial team will be called an expert SLP. The expert SLPs that participated had not treated any of the children included as speakers in the studies. In study IV three professional listeners who were not experts in cleft palate speech also participated. These will be referred to as the non-expert SLPs in this text even though they might be experts in another area of speech and language pathology. Untrained listeners participated in studies II and IV, 26 of the untrained listeners were working and did not have a profession working with children on a daily basis or an educational background in linguistics, the other six untrained listeners were first year undergraduate SLP students. They were chosen to represent the general adult public which a speaker will encounter in their daily life. Most of the untrained listeners had a background in the northern region of Sweden (the same area as the speakers in our study). According to self-report they had normal hearing. See table 2 for number and type of listeners included in each study.

Table 2. Description of listeners

<table>
<thead>
<tr>
<th>Group</th>
<th>n (F, M)</th>
<th>Age range</th>
<th>Study</th>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert SLPs I</td>
<td>2 (2F)</td>
<td>40-45</td>
<td>I, II</td>
<td>At least four years on cleft palate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>team</td>
</tr>
<tr>
<td>Expert SLPs II*</td>
<td>3 (3F)</td>
<td>45-55</td>
<td>IV</td>
<td>At least nine years on cleft palate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>team</td>
</tr>
<tr>
<td>Non expert SLP</td>
<td>3 (3F)</td>
<td>28-49</td>
<td>IV</td>
<td>At least two years of clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>experience</td>
</tr>
<tr>
<td>Untrained gp I</td>
<td>28 (12M, 16F)</td>
<td>20-41 yrs</td>
<td>II</td>
<td></td>
</tr>
<tr>
<td>Untrained gp II</td>
<td>6 (0M, 6F)</td>
<td>20-49 yrs</td>
<td>IV</td>
<td></td>
</tr>
</tbody>
</table>

* The two members of Expert SLPs I were also members of Expert SLPs II

Assessment materials and instrumentation

Speech stimuli

Three sets of sentences were used for auditory perceptual analysis (appendix 1) and three sets for recording with the nasometer (appendix 2). There is an overlap for a set of sentences with oral only speech sounds. This was used for both the perceptual assessments and the nasometry recordings. For an overview of speech stimuli for each study see table 3.
### Table 3. Speech stimuli for auditory perceptual tasks and nasalance assessments

<table>
<thead>
<tr>
<th>Stimulus description</th>
<th>Characteristics</th>
<th>Purpose</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence sets A &amp; B*</td>
<td>Contain</td>
<td>Auditory perceptual assessment</td>
<td>I and II</td>
</tr>
<tr>
<td></td>
<td>- oral only material</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- high pressure consonants</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- adjacent nasal consonants and pressure consonants</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>- varied manner and place of articulation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Auditory perceptual assessment</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nasalance measures for oral only speech stimulus</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Set of oral only sentences</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10 sentences selected from SVANTE test</td>
<td></td>
<td>III, IV</td>
</tr>
<tr>
<td></td>
<td>Set of mixed phonetic content sentences</td>
<td>Nasalance measures</td>
<td>III, IV</td>
</tr>
<tr>
<td></td>
<td>7 phonetically balanced Swedish sentences</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Set of mixed phonetic content loaded with nasal phonemes (so called nasal sentences)</td>
<td>Nasalance norms</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td>7 nasally loaded sentences</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oral words</td>
<td>Nasalance norms</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td>Spontaneous speech</td>
<td>Auditory perceptual assessment</td>
<td>IV</td>
</tr>
<tr>
<td></td>
<td>Elicited with pictures or retell</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Used in Swedish cleft palate clinics before the introduction of SVANTE

---

**Audio recordings of perceptual speech stimuli**

Audio tape recording has been made using either a Foster D-5 digital master recorder and a Sennheiser microphone extended from the ceiling in a treatment room or in sound proof recording studio using Panasonic SV-3800 digital audio tape recorder and microphone AKG C 420 mounted on a head set 10/15 cm from the mouth. The equipment in the recording studio is calibrated so that the microphone distance of 10 (earlier recordings) or 15 cm (later recordings) has been taken into account.

All sentences material was elicited by repetition, both for readers and for non-readers to get comparable materials. Spontaneous speech was elicited by a picture for non-readers and by retelling a story by readers. CDs for perceptual assessment were constructed by editing and randomisation of recordings, between 23 and 35% of recordings were repeated for evaluation of intra-rater reliability.
Assessment form for expert and non-expert SLPs

The assessment form used for auditory perceptual analysis in studies I and II contained the variables: hypernasality, hyponasality, audible nasal emission and/or nasal turbulence, weak pressure consonants and articulation variables, but the articulation variables are not reported in this thesis (appendix 3). A second version without the variable weak pressure consonants and the articulation variables was used for study IV. All variables were rated on a five point ordinal scale, see table 4. In study IV a slightly more detailed description of the scalar points were used, see appendix 4.

Table 4. Description of scalar points for expert raters.

<table>
<thead>
<tr>
<th>Scale value</th>
<th>Hypernasality, hyponasality, audible nasal air emission and/or nasal turbulence</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal speech</td>
</tr>
<tr>
<td>1</td>
<td>Slight deviation*/ Single occurrence</td>
</tr>
<tr>
<td>2</td>
<td>Mild deviation/ Some occurrences</td>
</tr>
<tr>
<td>3</td>
<td>Moderate deviation/ Frequently occurring</td>
</tr>
<tr>
<td>4</td>
<td>Severe deviation/ Occurring always or close to always</td>
</tr>
</tbody>
</table>

* This is a very slight deviation, just to indicate that speech is not completely without nasality.

Assessment form for untrained listeners

For this project an assessment form for untrained listeners was developed from a pilot version used in an earlier study (Fransson, 2002). The aim was to develop a description of the variables that were equivalent to the ones on the expert assessment form but in every day language. Results of investigations within the present study confirmed that untrained listeners did not differentiate between types of nasality i.e. “speaking through the nose” and “has a blocked nose”. Thus, on the untrained listeners form this was reduced to one variable “speaking through the nose/has a blocked nose”. The variable “puffs of air coming from the nose” was used to describe audible nasal emission and/or nasal turbulence and the articulation variable was described as “This child has other difficulties in pronunciation”. The form also contained the statement “This child needs help with his/her speech, e.g. speech therapy” as a binary question with yes and no as options in order to capture the subjective opinion on how severe the untrained listener thought the speech disorder was. This version of the form was used for study II, see appendix 5. For study IV the variable about articulation was removed and the wording of scale value descriptors was changed slightly so that in study IV all listeners had the same description of scalar points, see appendix 6.
Experimental procedures

Auditory perceptual assessment

Expert and non-expert SLPs: All professional listeners used headphones for listening and filled out the assessment form independently. For study I and II there was also a consensus rating by the two experts, they first rated each speaker independently and then in case of disagreement reached a consensus score. Assessments were made from CDs with headphones in quiet surroundings. The listeners were allowed breaks during the listening task.

Untrained listeners: Untrained listeners listened to recordings and filled out the assessment form independently. Seven untrained listeners listened simultaneously through high quality loudspeakers in a quiet teaching room, twenty five listened through high quality headphones in a quiet room and two listeners did part of the listening through head-phones and part through loud speakers due to technical problems.

Measurement of nasalance

Nasometer recordings (Studies III, IV) were made with a Kay Pentax Nasometer™II, model 6400 (Kay Elemetrics, 2001) and a laptop computer (Dell Latitude). The laptop computer and the Nasometer™ were connected to a computer docking station to accommodate requirements for the Nasometer's sound recording card. At the time of purchase it was not possible to use a laptop computer without a docking station. The head set was placed according to instructions in the manual.

All sentence material was elicited by repetition, both for readers and for non-readers to get comparable materials.

Test-retest reliability was obtained from repeated recordings of 12% of children for the whole speech stimulus in study III, and from all children on one speech stimulus each (oral, mixed or nasal) in study IV. For both studies the repeated recordings were made in the same session with no headset removal.

Data analysis

Reliability of auditory perceptual ratings for expert SLPs

Inter-rater reliability was calculated using all judgements excluding repeated ratings on the same speech stimulus. In studies I and IV the reliability was calculated with point-to-point agreement and ± one scale value, and also with weighted kappa ($\kappa_w$). In study II Spearman’s correlation was used in conjunction with $\kappa_w$. $\kappa_w$ is a version of Cohen’s kappa that weights ratings in order to give two ratings with one scale point distance a higher weight than ratings with larger distance between scale points. It is a conservative measure because any agreement that could have been by chance is assumed to have been obtained by chance (Cordes, 1994). Values of $\kappa_w$ can be interpreted as follows: 0.41-0.60 moderate agreement, 0.61-0.80 good agreement, >0.80 very good agreement (Altman, 1991).

Intra-rater reliability was conducted for all listening tasks in studies I, II and IV using the duplicate recordings (30-35%) that were incorporated into the randomised
samples contained on the listening CDs. Intra-rater judgements on these duplicate recordings were compared using point-to-point agreement (studies I, II, and IV) and ± one scale value (study IV).

An additional investigation on inter-rater reliability has also been conducted in this thesis by combining data from studies I and IV for the two expert SLPs who were participants in all three studies. Data from auditory perceptual ratings for expert SLPs 1 and 2 have been presented in cross-tabulations to extend the individual investigations of their inter-rater reliability. In the cross-tabulations the scale values 0 and 1 has been combined since they represent normal resonance and a slight change in resonance, very close to normal. See table 5.

<table>
<thead>
<tr>
<th>Study</th>
<th>Listeners</th>
<th>Listening material</th>
<th>Number of speakers</th>
<th>Inter-rater measures</th>
<th>Intra-rater measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Two experts (A, B)</td>
<td>Sentences A and B</td>
<td>38</td>
<td>( \kappa_w ), point-to-point, +/- 1, cross-tab</td>
<td>point-to-point</td>
</tr>
<tr>
<td>II</td>
<td>Two experts (A, B)</td>
<td>Sentences A and B</td>
<td>36</td>
<td>( \kappa_w, \tau_s ), ranking</td>
<td>point-to-point</td>
</tr>
<tr>
<td>IV</td>
<td>Three experts (A, B, C)</td>
<td>Sentences C, spontaneous speech</td>
<td>25</td>
<td>( \kappa_w ), point-to-point, +/- 1, ranking</td>
<td>point-to-point, +/- 1</td>
</tr>
</tbody>
</table>

**Comparison between expert SLPs and untrained listeners**

Comparisons were made using ratings of the resonance variables the ‘audible nasal emission’/‘puffs of air’ variables and the ‘need for intervention’ question.

The listener ratings of all speakers from studies II and IV have been used to compare expert SLPs and untrained listener ratings. For the expert SLPs two alternate types of ratings were used as comparison with untrained listeners. For the data from study II, a consensus rating from two expert SLPs was obtained, while from study IV a median rating from three expert SLPs was calculated. For untrained listener ratings it was always median scores for each speaker by the untrained listener groups that were used (28 listeners in study II and 6 in study IV).

Comparisons of ratings between the two groups were made using Spearman’s correlation and by a qualitative comparison of the level of agreement between median/consensus expert ratings and median untrained listener ratings for each speaker.

**Swedish norms for the Nasometer™**

In the norms study mean and standard deviation (SD) were calculated for each type of speech stimulus. Significant differences between groups on nasalance scores for regional dialect, gender and age were investigated by a three-way-analysis of variance (ANOVA). Post-hoc testing with student’s t-test was performed where differences were found. A correlation analysis with Pearson’s correlation coefficient was also performed to assess the correlation between the various types of speech stimuli. Test-retest reliability was calculated on 12% of the participants who were recorded twice within the same recording session. Retest reliability was measured as the difference between first and second recording for each speech stimulus.
Comparison between perceptual ratings and nasalance scores

Perceptual ratings from three listener groups (expert SLPs, non-expert SLPs and untrained) were compared with nasalance scores. Comparisons were made in two ways – a correlation analysis and a qualitative analysis to establish a nasalance threshold for the presence of hypernasality.

Correlations were derived between group medians of auditory perceptual ratings of the nasality variable for each speaker on two types of stimulus (oral sentences and spontaneous speech) and nasalance scores on both oral sentences and mixed (phonetically balanced) sentences. A measure of non-parametric correlation, Spearman’s rank correlation ($r_s$), was used to correlate auditory perceptual ratings of hypernasality with nasalance scores. According to Colton (Colton, 1974) correlations between 0.50 and 0.75 show a moderate to good relationship, and those greater than 0.75 a very good to excellent relationship. Statistical differences between correlations for the various listener groups and stimulus types were calculated using Kruskal Wallis one way analysis of variance followed by post hoc Mann Whitney U tests (for effect of listener group), and Wilcoxon Signed rank tests for effects of stimulus type.

The establishment of a nasalance threshold involved ranking speakers according to nasalance score on oral sentences and inspecting the corresponding perceptual ratings to obtain an optimal score below which speakers were not judged to have hypernasality present in their speech.
RESULTS

Reliability of auditory perceptual ratings for expert SLPs

Inter-rater reliability

Inter-rater reliability on hypernasality for expert SLPs in our studies show values between 45 and 60% for exact point-to-point agreement, when calculated within± one scale value the agreement was between 84 and 96%, the weighted kappa was $\kappa_w$ 0.45-0.64. For hyponasality exact agreement was between 80 and 100%, within ± one scale value the agreement was 80-100%, the weighted kappa was $\kappa_w$ 0.61-0.62. For weak pressure consonants the exact agreement was 63%, within one scale value the agreement was 92% and $\kappa_w$ 0.43-0.48. For audible nasal air emission/nasal turbulence the exact agreement was 60-84%, within one scale value was 92-96% of agreements, and the was $\kappa_w$ 0.54-0.83. See table 6.

Table 6. Inter-rater reliability for expert SLPs. Figures from three studies (I, II, IV). 1,2

<table>
<thead>
<tr>
<th></th>
<th>Study I 3</th>
<th>Study II</th>
<th>Study IV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>+/-1  $\kappa_w$</td>
<td>r_s $\kappa_w$</td>
</tr>
<tr>
<td>Hypernasality</td>
<td>45</td>
<td>84.05* 0.48</td>
<td>48-60 84-96 0.49-0.64 0.55</td>
</tr>
<tr>
<td>Hyponasality</td>
<td>82</td>
<td>97.60* 0.61</td>
<td>80-100 100 - -</td>
</tr>
<tr>
<td>Aud. nasal air emission/nasal turbulence</td>
<td>66</td>
<td>92.86* 0.76</td>
<td>60-84 92-96 0.54-0.83 0.69</td>
</tr>
</tbody>
</table>

1 In a few instances the figures can not be found in the corresponding article but have been calculated specifically for this summary of the studies.
2 The same two experts in study I & II. Joined by a third expert in study IV.
3 Only calculated for speakers with CLP/CPO.
* p<.001

Cross-tabulations show the distribution of ratings for hypernasality and audible nasal air emission/nasal turbulence for the two expert raters that have rated all 52 speakers with CLP, CP or velopharyngeal impairment in study I and IV. The cross-tabulation was made to show instances were one rater has rated normal-mild deviation when the other rater has rated moderate-severe deviation which would be highly relevant in clinical decision making. Seven out of 52 ratings of hypernasality, 13%, show a disagreement of this kind and three out of 52, 6%, ratings for audible nasal air emission/ nasal turbulence. See table 7.
Table 7. Number of ratings for each scale value (scale with four scalar points: 0-1, 2, 3, 4) for two expert SLPs on hypernasality and audible nasal air emission/nasal turbulence. Only ratings for patient groups have been included.*

**Hypernasality**

<table>
<thead>
<tr>
<th>Scale Value</th>
<th>Rater 1</th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>21</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>27</td>
<td>15</td>
<td>9</td>
<td>1</td>
<td>52</td>
</tr>
</tbody>
</table>

**Audible nasal air emission/nasal turbulence**

<table>
<thead>
<tr>
<th>Scale Value</th>
<th>Rater 1</th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>20</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>15</td>
<td>13</td>
<td>2</td>
<td>52</td>
</tr>
</tbody>
</table>

* Bold numbers indicate where one rater has rated normal/mild and the other rater has rated moderate/severe which is relevant for clinical decision making. Ratings from study I (n=38) and study II (n=14), auditory perceptual ratings on sentences.

**Intra-rater reliability**

Intra-rater reliability for expert SLPs in our studies has been calculated. For hypernasality the exact point-to-point agreement was between 55 and 88%, the $\kappa_w$ was between 0.17 and 0.58. For hyponasality exact agreement between 73 and 100%. For audible nasal air emission/nasal turbulence the exact agreement was between 75 and 100%, kappa values $\kappa_w$ 0.73-0.84. See table 8.

Table 8. Intra-rater reliability for expert raters. Figures from three studies (I, II, IV)

<table>
<thead>
<tr>
<th>Study I</th>
<th>Study II</th>
<th>Study IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>Con-sensus</td>
</tr>
<tr>
<td>% / $\kappa_w$</td>
<td>% / $\kappa_w$</td>
<td>% / $\kappa_w$</td>
</tr>
<tr>
<td>Hypernasality</td>
<td>79</td>
<td>0.58</td>
</tr>
<tr>
<td>Hyponasality</td>
<td>100</td>
<td>*</td>
</tr>
<tr>
<td>Aud. nasal air emission/nasal turbulence</td>
<td>86</td>
<td>0.73</td>
</tr>
</tbody>
</table>

* Not frequent enough to allow for calculation of weighted kappa

**Comparisons between expert SLPs and untrained listeners**

**Inter-group correlations**

In study II the correlation for rating of resonance between the median for expert speakers and median for untrained listeners was $r_s$ 0.62, $p<0.001$. In study IV the correlation between median rating of experts and median rating for untrained listeners was $r_s$ 0.80, $p<0.001$, for sentences and $r_s$ 0.88, $p<0.001$, for spontaneous speech.
Qualitative comparison of ratings

A combination of findings from studies II and IV indicate that altogether untrained listeners found 19 speakers with a median rating of $\geq 2$ on resonance (11 in study II and 8 in study IV) which were rated as follows by the expert SLPs: 10 moderate-severe hypernasality, 2 moderate hyponasality, 3 mild hypernasality, 1 slight hypernasality, 1 no nasality, 2 with no nasality but comments about other voice characteristics. Three speakers were rated to have mild hypernasal resonance by expert SLPs but had a median lower than 2 for ratings by untrained listeners. Table 9 shows cross-tables of median/consensus ratings by expert SLPs and untrained listeners for all speakers in study II and IV on resonance and audible nasal emission or ‘puffs of air’. These cross-tables show how well the ratings agree; ratings have been divided into normal (0-1) and hypernasal (2-4).

The speakers in study II and IV that had a rating of 3 or 4 (moderate or severe) (n=10) on hypernasality by expert SLPs had by the untrained listeners a range of 67-100% ‘yes’ for the statement ‘Does this child need help with his/her speech’. The six speakers with a rating of 2 for hypernasality (mild) had the following percentages of ‘yes’: 0, 16, 18, 50, 50, 100%.

In both studies (II, IV) it was found that untrained listeners noted audible nasal emission but to a much lesser extent than the expert SLPs did. In study II the expert SLPs found 13 speakers with nasal emission (rating of $\geq 2$) and untrained listeners found 6 speakers, while in study IV the expert SLPs found 7 speakers and the untrained listeners found 2 speakers in the sentence rating task and the expert SLPs found 3 speakers, and the untrained listeners found 3 speakers in the spontaneous speech rating task.

Results from study II and IV show that both expert and untrained listeners are good at separating out the reference speakers who did not receive median ratings higher than 0 or 1 on any variable. Even the individual ratings to a very large extent consist of ratings of 0 or 1, with only occasional ratings of 2 by individual untrained listeners and no ratings of 3 or 4.
Table 9. Cross-table of resonance and audible nasal air emission/nasal turbulence ratings for expert and untrained listeners for all speakers in study II and IV. Consensus/median for experts, median for untrained listeners.*

<table>
<thead>
<tr>
<th>Resonance</th>
<th>Ratings untrained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-1</td>
</tr>
<tr>
<td>Ratings</td>
<td>0-1</td>
</tr>
<tr>
<td>experts</td>
<td>2-4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audible nasal air emission</th>
<th>Ratings untrained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0-1</td>
</tr>
<tr>
<td>Ratings</td>
<td>0-1</td>
</tr>
<tr>
<td>experts</td>
<td>2-4</td>
</tr>
</tbody>
</table>

* from study IV ratings from perceptual ratings of sentences have been used

**Swedish norms for the Nasometer™**

A single mean score was calculated for the whole group of school aged children: oral sentences 12.7 (5.6) %, mixed sentences 29.5 (6.1)%, nasal sentences 56.5 (6.4) %. See table 10 for mean, SD and mean+2SD for oral and mixed sentences for school aged children.

Table 10. Means and standard deviations of nasalance scores for all oral and mixed sentence speech stimuli for school children and pre-school children respectively.

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Age</th>
<th>n</th>
<th>Oral</th>
<th>Mean+2SD</th>
<th>Mixed</th>
<th>Mean+2SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>School age</td>
<td>6-7 &amp; 9-11</td>
<td>175</td>
<td>12.7 (5.6)</td>
<td>23.9</td>
<td>29.5 (6.1)</td>
<td>41.7</td>
</tr>
</tbody>
</table>

There were no significant differences due to regional dialect or gender. For age there was a significant difference on nasal sentences between the youngest children and the two older groups, age 4-5 vs. age 6-7 (t=-2.844, p=0.006) and for age 4-5 vs. age 9-11 (t=-2.888, p=0.005).

Test-retest values are found in table 11.

Table 11. Test-retest reliability for nasalance scores

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>% of retest scores within ±3</th>
<th>retest scores within ±5</th>
<th>test-retest correlation (r)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oral words</td>
<td>80.5</td>
<td>96.3</td>
<td>.900</td>
</tr>
<tr>
<td>Oral sentences</td>
<td>92.9</td>
<td>100</td>
<td>.930</td>
</tr>
<tr>
<td>Mixed sentences</td>
<td>71.4</td>
<td>96.4</td>
<td>.878</td>
</tr>
<tr>
<td>Nasal sentences</td>
<td>78.6</td>
<td>85.7</td>
<td>.923</td>
</tr>
</tbody>
</table>

*All r significant at p=0.01.

**Comparison between perceptual ratings and nasalance scores**

**Correlational analysis**

There was a significant correlation between perceptual ratings and nasalance scores for all three listener groups (expert SLPs, non-expert SLPs, untrained listeners). Correlations for expert SLPs was $r_s 0.67-0.74$ (good), mean 0.70, for non-expert SLPs $r_s 0.55-0.76$ (moderate-good very good), mean 0.66, for untrained listeners 0.42-0.56 (fair-moderate), mean 0.48. All correlations were significant, for both groups of SLPs at the p-level 0.01, for untrained listeners for p-level 0.05. The expert SLPs'
Correlations between perceptual ratings and nasalance scores (0.70) were significantly higher than the untrained group’s correlations (0.48), \((U=31.5, p<0.001)\). The difference between correlations for non-expert listeners and untrained listeners was not significant at \(p<0.01\).

Correlations between perceptual ratings on spontaneous speech and oral nasalance scores where higher (mean 0.69) than correlations between perceptual ratings of oral stimulus and oral nasalance scores (mean 0.58), this was however not significant \((z=-1.098, p=0.272)\).

There was a significant difference for correlations between perceptual ratings and oral nasalance scores (mean 0.51) as compared to correlations between perceptual ratings and mixed nasalance scores (mean 0.47). Correlations between perceptual ratings and nasalance scores was significantly higher when oral nasalance scores were used \((z=-3.572, p<0.001)\).

**Nasalance threshold for presence of perceived hypernasality**

In order to find a nasalance threshold for normal resonance contra hypernasality the ratings of expert SLPs and non-expert SLPs was investigated. An inspection of the results indicates a nasalance threshold in the vicinity of 26-29% for the oral stimulus. Table 12 shows that there was only one speaker with nasalance score \(\leq 26\) who were perceived to have hypernasality and four with nasalance score \(>26\) who were perceived to have normal resonance. Only speakers from the patient group received nasalance scores \(>26\) i.e. (range 29-65%) all speakers from the reference group had nasalance scores \(\leq 26\) (range 6-26%). It was not possible to establish a nasalance threshold using the untrained listeners’ ratings.

<table>
<thead>
<tr>
<th>Nasalance</th>
<th>Expert and non-expert SLPs’ ratings of hypernasality (oral sentences)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\leq 26)</td>
<td>Absent/ within normal limits</td>
<td>26</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(&gt;26)</td>
<td>Present</td>
<td>3</td>
<td>1</td>
<td>11</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 12.** Cross-table comparing expert SLPs’ ratings with nasalance scores.
DISCUSSION

Performing auditory perceptual assessments of nasality, especially hypernasality, is potentially difficult. As a starting point for the research within this doctoral project an outcome study of speech in children with cleft palate was conducted. Care was taken to use proper methodology such as: listeners who are not biased through working with the children, good listening conditions, recorded speech samples, consensus listening etc. Even in this case it was found that inter- and intra-rater reliability was poor (only acceptable). In order to increase the understanding of hypernasality and the assessment there of, different aspects of evaluation of nasal speech were investigated further.

Reliability of auditory perceptual ratings for expert SLPs

Issues of rater reliability in assessment of resonance disorders have been discussed in the literature and it has become clear that perceptual rating of hypernasality is a challenging task. The difficulties have become evident through low rater reliability figures (Counihan & Cullinan, 1970, Keuning et al., 1999, Kuehn et al., 2002; Persson et al., 2006). To add to the problem a review by Lohmander and Ohlsson (2004) has pointed out that many articles do not report intra- and/or inter-rater reliability which makes it impossible to make any comparison with many studies on relevant topics. A review of recent articles found with the search string “cleft palate speech” in the database PubMed indicates that there is still a lack of data on reliability (e.g. Isotalo, Pulkkinen, & Haapanen, 2007; Merrick, Kunjur, Watts, & Markus, 2007; Khosla, Mabry, & Castiglione, 2008; Andrades et al., 2008; Mink van der Molen, Janssen, Specken, & Stubenitsky, 2008; Priester & Goorhuis-Brouwer, 2008).

In the current studies the expert SLPs had moderate to good inter-rater reliability (κw 0.45-0.64) and poor exact point-to-point agreement between 44.7% and 60% (within± one scale value between 84% and 96%) on the variable hypernasality. On ratings of hyponasality and audible nasal air emission/nasal turbulence the weighted kappa values for inter-rater reliability indicated moderate–very good agreement but generally good agreement. Thus the results from this project indicate that inter-rater reliability was lowest for hypernasality, implying that it was the most difficult variable to assess. Intra-rater agreement for hypernasality in the current studies measured point-to-point was between 46% and 100% and within one scale value 81-100%.

Other studies have reported higher reliability figures for perceptual assessment of hypernasality by expert listeners. Lewis et al. (2003) reported weighted kappa values between 0.71 and 0.73 for three expert raters. In a multi-center study by Grunwell et al. (2000) a number of expert raters reached complete agreement (alpha-coefficient of estimated rank order) on hypernasality. Pulkkinen et al. (2001) reached 91-94% exact agreement on various variables including hypernasality. Sell et al. (2001) reached high kappa values for hypernasality (κ=0.81) but lower on audible nasal emission (κ=0.59). These examples show that it is possible to reach high values on inter-rater reliability, but there are differences in task design that may explain the discrepancy between the reliability found in the current studies and those with higher reported reliability. Two studies, Grunwell et al. and Pulkkinen et al. had fewer than 5 scale points (3 and 2 respectively), while another (Sell et al.) included extensive training for the listeners prior to the study. However, another study that employed
listener training did not reach more than acceptable inter-rater reliability (Kuehn et al., 2002). In the study by Lewis et al. (2003) listener ratings was based on listening to only one sentence. Ratings from such a short stimuli could diminish the risk that the listeners attended to different parts of the stimulus (Watterson, Lewis, Allord, Sulprizio, & O'Neill, 2007). Other factors that have been reported to affect inter-rater reliability in assessment of hypernasality are the quality of recordings, description of the scalar points, the number of variables that are simultaneously assessed and the use of anchor speech samples. In study IV each of these factors was attended to, in an effort to improve listener reliability.

The quality of recordings was ensured by making recordings under standardised conditions (Asplund, 2004). Listeners used good quality head phones in order to ensure good listening conditions. The scalar points were a little more defined in study IV than in study I but could probably be even more defined. In addition, in study IV the task involved rating of less variables than in study I and II with only three speech variables to rate: hyper- and hyponasalisation and audible nasal air emission/nasal turbulence. The speech samples included few speakers with articulatory disorders, that have been reported to interfere with assessment of nasality (McWilliams, 1990, Kent, 1996). In planning the last study (IV) a reference-CD was introduced, with so called anchor stimuli, with examples of four degrees of hypernasality in order to calibrate the listeners and thereby hopefully improve inter- and intra-rater reliability (Gerratt, Kreiman, Antonanzas-Barroso, & Berke, 1993). Listeners heard this CD four times during the rating of the voices. However these improvements in task introduced into study IV did not result in remarkably higher kappa values for inter-rater reliability for expert SLPs. The reliability went from 0.45 and 0.48 (in study I and II respectively) to 0.49-0.64 (mean 0.55 ) for study IV. We can thus not say that anchor stimuli improved reliability in this study. It should be noted, however, that in study IV there were three different orders of presentation of stimulus voices to avoid order effects, which would likely lower the reliability since ratings are influenced by recollection of previous voices.

There are more examples of good inter- and intra-rater agreement with assessment with the British and Irish assessment material CAPS-A (John et al., 2006, Sell et al., 2008). This seems to have been achieved by careful development of an instrument (the CAPS-A) and rater training specifically for that instrument, where at least one day of training has been included in the reported studies. In John et al. (2006) it is described how the inter- and intra reliability was not satisfactory even after training in their pilot study and they therefore worked more on the descriptions of the scalar points and achieved higher reliability measures in a second pilot study (inter-rater reliability for hypernasality ICC 0.88 across seven listeners and intra-rater reliability κ 0.62 for all listeners). Examples of more detailed descriptions of scalar points are “hypernasality is evident on close vowels” for mild hypernasality and “there is increased nasality heard on high and low vowels” (Henningsson et al., 2008, John et al., 2006). This type of description is more detailed than is commonly found in the literature. The description of scalar points for hypernasality in CAPS-A (John et al., 2006) does not include any ratings of consistency. In the present study the scale descriptions include both degree and consistency in the same scale. They have also reduced the number of scalar points for hyponasality and nasal air emission/ nasal turbulence. In the study by Sell et al. (2008) the rater reliability was high for intra-rater measure but not as high for inter-rater measure for 36 raters (ICC 0.45-0.50). The listeners had an option to use a rating of ‘8’ to indicate that they considered there
was not sufficient speech to rate the speaker on a certain variable. The authors note that the rating of ‘8’ was most commonly used for ‘hypernasality’ which is another indication that the variable is difficult to rate.

According to two previous studies (Zraick & Liss, 2000; Whitehill, Lee, & Chun, 2002) direct magnitude estimation would be better than the commonly used equal appearing interval (EAI) scales for measurement of hypernasality. The DME scale where the listener estimates a value to a sample in comparisons to an anchor sample fits our perception of nasality better and give higher reliability of ratings. Even though DME seems more appropriate for ratings of nasality than EAI-scales, it is not likely that DME will be used in clinical settings since it is not practical and hard to communicate to patients and colleagues (Whitehill et al., 2002). Zraick et al. (2000) did conclude that there was not enough information to rule out the EAI-scales for assessment of hypernasality. The reason an ordinal scale was used in our study was that it is the one in current clinical use and has not been criticised like the EAI-scales. The ordinal scale used in this study had more narrow gaps between the values at the lower end of the scale.

In light of the results in the current studies and in the literature it seems there are several aspects of perceptual assessment of nasality that warrant further study; specific listener training and the impact of anchor stimuli, the impact of type and content of scales as well as impact of speech stimulus.

**Clinical implications**

The reliability scores for ratings by the two SLPs who participated in all perceptual assessment were re-examined in order to investigate the relative importance of inconsistencies between raters. The ratings were entered into cross-table comparison where one can see that 13.5% of ratings of hypernasality fell on opposite sides on the mild/ moderate border of ratings. This is considered interesting since this is of clinical importance: most of the time a rating of “moderate” is likely to lead to further assessment and suggestions for intervention while a “mild” rating is usually not a reason for further assessment unless the patient or the parents express that speech resonance is a problem to the patient in daily life. In ratings of audible nasal air emission/nasal turbulence there was 5.8% that had rating on opposite sides of the border between mild and moderate.

The implication of low inter-rater reliability is that a patient may get different treatment decisions depending on who makes the evaluation, while the implication of low intra-rater reliability is that the same speaker may get different decisions from the same SLP on different occasions. In the clinic a treatment decision for surgical treatment is not based on one visit alone, the patient will first be evaluated at least on two occasions and often with two SLPs involved in the decision making, especially for those cases that are perceived as border line. The largest problem with low reliability is probably not in the clinical situation but in research aimed at comparing treatment methods and outcomes. One has to remember that the tasks in a research study are conducted under quite different situations from the tasks in a clinical situation (Gooch, Hardin-Jones, Chapman, Trost-Cardamone, & Sussman, 2001). In research it is important to use recordings instead of live assessments or information from patient records to avoid the influence of expectations, visual information and spoken information from the patient (Podol & Salvia, 1976; Mehendale & Sommerlad, 2003). However, in a clinical situation information from the formal testing is complemented
with and potentially biased by informal assessment during the visit, reports from patient and parents, both audio recording and live impression of the speech, several types of speech samples from naming, repetition of sentences as well as continuous speech. Another aspect in the clinical decision about further investigation and intervention is that it usually does not rely solely on the assessment of hypernasality. The variables of audible nasal air leakage/nasal turbulence and weak pressure consonants are also weighted in the decision making and for most patients with CLP hypernasality usually co-exists with either or both of the other variables. Occasionally, in some patients with VPI without CLP hypernasality is found in the absence of other speech symptoms of VPI but mostly co-occurs with either audible nasal air leakage/nasal turbulence or weak pressure consonants. Listener training also has a role in clinical practice. Team SLPs should have listening training as part of their continuing education.

It is still a concern that so much difference is found in expert SLP assessments. In the interest of finding optimal treatment, e.g. surgical protocols for cleft palate surgery, it is important to have reliable measures of speech variables to allow comparisons. One option in addition to working on better reliability of perceptual assessments is to complement listening with instrumental measurements which is already implemented in many clinics, especially acoustic and air flow techniques. See more below for the use of the Nasometer as a complement to auditory perceptual assessments.

**Comparisons between experts and untrained listeners**

The purpose of comparing untrained listeners ratings’ with those of expert SLPs was to investigate if people that speakers with nasal speech encounter in everyday situations notice hypernasality and audible nasal air emission/nasal turbulence and consider this to be an indication of need for treatment. Our findings indicate that untrained listeners are aware of hypernasality in speakers, but are much less aware of audible nasal air emission/nasal turbulence, which is in line with clinical experience but no other records of this have been found in the literature. Earlier studies have both indicated that SLPs and untrained listeners give similar judgements (Starr et al., 1984; Tonz et al., 2002) and that untrained listeners note more hypernasality than expert listeners (Lewis et al, 2003). In Swedish clinics today a rating of moderate-severe hypernasality most often leads to further investigation and suggestions for treatment, while a mild hypernasality would normally not lead to further investigation or treatment. It is therefore of interest to see how ratings by expert SLPs and listeners compare and if speakers with mild hypernasality (as rated by expert SLPs) are perceived as hypernasal and in need of help by untrained listeners.

According to results from the current study there is no case for stating that experts in assessment of resonance disorders are overly sensitive and there is thus no risk for over-treatment of patients. Tönz et al. (2002) looked at which speakers were rated to have improved speech after surgery and found that SLPs and untrained listeners to a large extent identified the same speakers. This is in line with the findings of the present study where it was found that experts and untrained listeners to a large extent identified the same speakers as hypernasal and in need of treatment. It was also found that for the untrained listeners to notice audible nasal air emission/nasal turbulence, it had to be prominent or frequent. This finding agrees with clinical experience.
Results from study II and IV show that usually a rating of moderate to severe hypernasality by the expert SLPs corresponded to a need for intervention indicated by the untrained listeners. There was also a very good agreement in judging which speakers had no speech disorder in that no speaker from the reference group was rated >1.

In this study a contextual factor (WHO, 2002) that may influence a speaker with hypernasality was investigated, namely whether the everyday listener notices the speech disorder and whether they think the speaker needs help with this. There are of course other contextual factors that may influence functioning and disability such as cultural beliefs of health and disabilities, communicative culture, acoustic properties of class rooms and work environment etc (WHO, 2002). This study does not examine how people with this type of speech disorder are perceived and whether hypernasality and audible nasal air emission/nasal turbulence limit their participation in daily life in different ways. However, other studies have indications that hypernasality or cleft palate speech is perceived negatively by untrained listeners (Blood & Hyman, 1977; McKinnon, Hess, & Landry, 1986; Lallh & Putnam Rochet, 2000). Since the present study shows that hypernasality is noticed by untrained listeners and other studies have found that there is a negative perception associated with hypernasality a tentative conclusion is that hypernasality may well impact on participation.

**Swedish norms for the Nasometer™**

The reason for conducting a study to establish norms for Swedish nasalance scores (study III) was because no such study had been conducted and there was a need for normative data before we conducted a study to compare auditory perceptual ratings and nasalance scores as was done in study IV. The norms that are now available from that study will be also be useful in speech clinics in Sweden.

In our study we found that the mean values for oral sentences and mixed sentences were similar to those reported for languages such as Flemish, Canadian French and Hungarian. For oral stimuli similar means have been reported for English, Finnish and Thai as well. For nasal stimuli the mean from the present study was in the mid-range of those reported.

One normative score for different types of speech material can be used for children between 6-10 years of age: oral sentences 12.7 (5.6) %, mixed sentences 29.5 (6.1)%, nasal sentences 56.5 (6.4) %. The figure for +2SD may be used as a tentative threshold for normal and abnormal resonance for oral and mixed material (Kummer, 2005). For Swedish the cut-off would be 23.9% for oral sentences and 41.7% for mixed sentences.

Our study did not reveal significant differences for Swedish regional dialects. This is consistent with studies of Canadian English (Kavanagh et al., 1994), Irish English (Sweeney et al., 2004) and Hungarian (Hirschberg et al., 2006) but different from a study on North American English which reported differences (Seaver et al., 1991). There was no gender difference in our study which is in line with findings in other Germanic languages (Nichols, 1999; Sweeney et al., 2004; van Doorn & Purcell, 1998) whereas studies from other branches on the language tree has found some differences due to gender (Prathanee et al., 2003; Hirschberg et al., 2006; Mishima et al., 2007).

For age there was a statistical difference between the youngest age group (5 year old) and the two older groups, but only on nasal sentences where the younger children
had a lower mean nasalance score (4 points). The only other study that has found differences due to age in a comparison between groups of children was a study of Thai (Prathanee et al., 2003). Otherwise a review of the literature indicates that there are only significant differences when children and adults are compared (Haapanen, 1991; Van Lierde et al., 2003). A study with a larger group of young children is needed to verify the finding of differences in nasalance score for nasal sentences.

The test-retest between two successive recordings for the same speaker in the present study found that 87% of the repeated recordings (all stimuli) were within four points difference and the rest were within eight nasalance points. This is a marginally higher than might have been expected in light of the fact the head set between successive recordings, but according to recent systematic studies of variability in nasalance scores our results are within limits for expected variation (Watterson et al., 2005; Lewis et al., 2008). Watterson et al. (2005) have found that intra speaker variation of 5 points is not uncommon in a group of speakers without speech disorders, and state the variation might be greater in a clinical population.

The clinical implications of this variability between recordings would be a recommendation to make more than one recording on the same speech stimulus in the same session, and to not interpret a change of 5 scale points between two single recordings as an improvement or deterioration in speech resonance unless it can be confirmed with repeated measures. Repeated recordings are also needed to confirm whether an individuals scores are inside or outside the statistical threshold for presence of hypernasality.

**Comparison between perceptual ratings and nasalance scores**

The findings in the study show a general trend for all three listener groups (expert SLPs, non-expert SLPs, untrained listeners) where higher perceptual ratings were correlated with higher nasalance scores. The correlations between listener ratings and nasalance scores from the Nasometer™ were good or moderate to good for SLP groups and fair to moderate for the untrained listeners. There was a significant listener group effect where expert SLP perceptual assessments had significantly higher correlations with the Nasometer scores than did the untrained listeners. The non-expert SLPs perceptual assessments also had higher correlations with the nasalance score than the untrained listeners’ but this difference was not significant. There are other studies supporting the finding of good correlations between expert SLPs perceptual ratings and nasalance scores (e.g. Dalston, 1991, Sweeney & Sell, 2008) but not support for the finding that expert SLPs had higher correlations than other groups (Watterson et al., 1996; Lewis et al., 2003; Keuning, Wieneke, & Dejonckere, 2004)

For nasalance scores on oral sentences there was no significant difference between correlations with perceptual ratings from spontaneous speech compared with perceptual ratings from oral sentences. This finding suggests that for clinical use perceptual ratings of oral material and spontaneous material work equally well in comparison with nasalance for oral sentences. There is thus no need to use exactly equivalent speech stimulus for perceptual ratings and nasometry as has been assumed in most research into the relationship between perceptual measures and nasometry (Watterson et al., 1993; Chun & Whitehill, 2001; Lewis & Watterson, 2003; Keuning et al., 2004)
The nasalance score for mixed sentences did not add any information and had statistically significant lower correlation with perceptual ratings than the score from the oral sentences. Thus a mixed stimulus is not necessary in a clinical evaluation of nasalance, a finding that is consistent with Dalston and Seaver (1992).

Besides the correlation analysis, qualitative examination of the data showed that three out of 25 speakers had conflicting results between perceptual assessments and nasalance scores, two speakers with low perceptual ratings (median of 0 or 1) and a fairly high nasalance score (48% and 45%) and one speaker with very low nasalance score (7%) but non-zero perceptual ratings of hypernasality. In clinical practice this would be the patient where it is of interest to note the discrepancy in the records and when necessary ask for a second opinion by a colleague and/or see the patient for a new assessment on a later occasion. This would of course depend on the patients opinion of his/her own speech and the need for further intervention. In the current study it was found that one of the speakers with conflicting results had received comments on the assessment form about the low quality of the recording which might have interfered with the perceptual ratings. The speaker with very low oral scores also had low scores on nasal sentences (not described in the current study), which may indicate nasal obstruction resulting in a change of resonance, maybe a mixed nasality.

Further qualitative examination of the data for the estimation of a nasalance score threshold for hypernasal vs non-nasal speech showed that for all speakers with a nasalance score of 26% or less on the oral stimulus the median rating of expert SLPs were 0 or 1. On the other hand, all speakers but two with a score of 29% or above had a median expert rating of 2 or above. This indicates a threshold between 26–29%, consistent with the statistically derived threshold of 24% (2SD above the mean) from the norms study.

We infer from the results that the Nasometer™ may be a valuable assessment tool in cases that are especially hard to judge and as a means to quantify hypernasality. In case of an SLP that is inexperienced in assessing resonance disorders the nasalance score will help in distinguishing between hyper- and hyponasal speakers.
CONCLUSIONS AND FUTURE RESEARCH

From the current studies it has been concluded that inter- and intra-rater reliability in ratings of speech nasality variables, particularly hypernasality, can be low for expert SLPs, even when good methodology has been applied. Importantly, there was a general agreement between expert SLPs and the untrained listeners in identifying which speakers are more hypernasal and in need of treatment. The untrained listeners were, however, less sensitive to audible nasal air emission/nasal turbulence than the expert SLPs. Furthermore, the expert speech-language pathologists’ auditory perceptual ratings and Nasometer™ scores correlated significantly better than the untrained listeners’ perceptual ratings and Nasometer™ scores.

Further study into what further measures can be taken to improve reliability of this important clinical and research tool is needed. Areas to further investigate are listener training and reference stimuli, modifying definitions of scalar points, and the use of other type of scales. Further study into the untrained listeners’ perception of hypernasal speech would be of great interest, both if and how they perceive hypernasality and investigating contextual factors, such as the attitudes towards this speech disorder.

Two interesting findings regarding the Nasometer™ score were found. Firstly, perceptual ratings from spontaneous speech may well be used in comparisons with nasalance scores from oral sentences. Secondly, the use of a nasalance score for mixed sentences does not add any extra information to the evaluation and does not need to be included. The good correlation between expert SLPs perceptual ratings and the nasalance scores implies that the Nasometer™ can be a good complement for listening in hard to judge cases, and can provide quantification of the resonance disorder to compare with the perceptual rating. Further, the results for the non-expert SLPs indicate that the NasometerTM may also assist the less experienced SLP making judgements especially the distinction between hyper-and hyponasality. Further study with a larger speaker group with a wide spread of speech nasality and larger listener groups of expert and non-expert SLPs is wanted to confirm the results of comparison between perceptual ratings and nasalance scores.

Eventhough there is low reliability in expert SLPs’ nasality ratings this does not seem to have clinical implications on treatment because different listener groups agree on what patients that need treatment. In addition, perceptual assessment by expert SLPs correlate well with Nasometer™ scores and those can be used to strengthen perceptual assessments for treatment decisions.

Normative values for Swedish in children aged 6-10 have been provided through the present study and can inform decisions in clinical settings in Sweden. One set of norms apply for both genders and at least three dialectal regions.
Acknowledgements

I would like to thank my supervisors for great support, quick reading and clear minds:

Professor Jan van Doorn, main supervisor, always with time for discussions and solving minor or major concerns.

Professor Anette Lohmander, supervisor, who started me on this line of research even before there was a plan for a PhD and with continuing enthusiasm throughout the project.

Many others have been helpful throughout the years, many thanks to all of you, especially: Christina Persson, Christina Havstam, Jill Nyberg for great co-operation.

All (34) untrained listeners who came here to listen to CDs, and Jocke, Monica, Anders and Ingrid who helped me to recruit these listeners.

All speakers who allowed me to make and use speech recordings.

Ulrika Fransson, Ida Bergh, Caroline Berg, Emelia Carlsson, former SLP-students at Umeå University, for contributions to this project. Professor Margit Tornéus for important input and cooperation.

Anders Asplund for support with editing of sound files and other essentials.

Three colleagues at Norrland University Hospital for devoting spare time to assess voices for this project. All my co-workers at the clinic and the division.

Students and teachers at schools in Umeå, Stockholm and Göteborg. Children and teachers at pre-schools in Umeå.

Josefin Enfält, my colleague, who has been a good sport when clinical work and writing a thesis didn’t match and for knowing that sometimes a bag of Dumle can make the whole difference. Members of the cleft palate team who showed interest and support all along.

The speech and language clinic, Västerbottens läns landsting for allowing me leave of absence for more than five years, especially the head of the clinic Käte Alrutz who has encouraged this research from the beginning.

Faculty of Medicine, Umeå University financing a PhD –student position.

Daniel Broman for discussing validity.

Christer for loving support, programming skills as well as finding me some untrained listeners.

My parents and my siblings with their families for always being there.
REFERENCES


Speech stimuli for auditory perceptual assessment  Appendix 1

**Version 1**
Pippis apa
Titta på TV
Kickis kakor
Koka ägg
En kopp kaffe
Emil är inte snäll
Solen lyser
Hämta inte mina vantar
Många mammor ammar

**Version 2**
Titta på TV
Kicki kokar ägg
I dag är det tisdag
Pippis apa hoppar
Ville å Valle
Sissi sover
Solen lyser
En kopp kaffe
Hämta inte mina vantar
Mamma och Mimmi är hemma

**Version 3**
Pippis apa piper
Bibbi bara jobbar
Titti tittar på TV
David å du leder
Kicki kokar korv
Giggi vill väga guld
Fiffi får kaffe
Vivvi vevar
Olle lurar Ella
Speech stimuli for Nasometer recordings

Appendix 2

Oral:
hus, filar, pippi, ljus, bilar, tv, gul, vit och duva

Pippis apa piper
Bibbi bara jobbar
Titti tittar på TV
David å du leder
Kicki kokar korv
Giggi vill väga guld
Sissi å Lasse sover
Fiffi får kaffe
Vivvi vevar
Olle lurar Ella

Oronasal:
Svante vill inte ha vantar
Eva har en vovve
Kalle har en kompis
Barnen går till skolan
Solen lyser
Finn tio fel
En kopp kaffe

Nasal:
Ankan simmar runt i vattnet
Kaninen äter morot
Anki hämtar hinken
Knut tändar lampan
Hämta inte mina vantar
Mormor hämtar tidningen
Mamma å Mimmi är hemma
## Assessment form expert SLPs, study I and II

### Appendix 3

**Voice_____________________ Rater _____________________**

<table>
<thead>
<tr>
<th>Disorder</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dysphonia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypernasality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyponasality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audible nasal air emission/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal turbulence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weak pressure consonants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Articulation**

- **Glottal articulation**
- **Pharyngeal fricatives**
- **Nasal fricatives**
- **Velar dentals**
- **Palatal dentals**

**Comments:** (note if you find the child has a speech or language disorder other than VPI/CLP-speech)

---

5 point scale

0. Normal speech
1. Slight deviation/ Single occurrence
2. Mild/ Some occurrences
3. Moderate/ Frequently occurring
4. Severe/ Occurring always or close to always
**Study IV, expert and non-expert SLPs’ assessment form**  
**Appendix 4**

**Speech assessment**

Circle the appropriate number

**Hypernasality**

<table>
<thead>
<tr>
<th>Description</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
</tr>
<tr>
<td>Slight or mild but intermittent</td>
<td>1</td>
</tr>
<tr>
<td>Mild but consistent</td>
<td>2</td>
</tr>
<tr>
<td>Moderate</td>
<td>3</td>
</tr>
<tr>
<td>Severe</td>
<td>4</td>
</tr>
</tbody>
</table>

**Hyponasality**

<table>
<thead>
<tr>
<th>Description</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
</tr>
<tr>
<td>Slight or mild but intermittent</td>
<td>1</td>
</tr>
<tr>
<td>Mild but consistent</td>
<td>2</td>
</tr>
<tr>
<td>Moderate</td>
<td>3</td>
</tr>
<tr>
<td>Severe</td>
<td>4</td>
</tr>
</tbody>
</table>

**Audible nasal air emission/nasal turbulence**

<table>
<thead>
<tr>
<th>Description</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Once or twice</td>
<td>1</td>
</tr>
<tr>
<td>Some times/several times</td>
<td>2</td>
</tr>
<tr>
<td>Often</td>
<td>3</td>
</tr>
<tr>
<td>Consistently or almost consistently</td>
<td>4</td>
</tr>
</tbody>
</table>

**Global evaluation**

This child needs help with his/her speech  
Yes  
No  
(Circle)

*Intermittent* = occurring every now and then
Read each statement and tick the appropriate box.

Child no: __________

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes, but very little</th>
<th>Yes, a little</th>
<th>Yes, quite a lot</th>
<th>Yes, a lot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. This child speaks through his/her nose or has a blocked nose</td>
<td>No</td>
<td>Yes, but only once or twice</td>
<td>Yes, sometimes</td>
<td>Yes, often.</td>
<td>Yes, almost all the time</td>
</tr>
<tr>
<td>2. It sounds like puffs of air coming through the nose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. This child has difficulties in pronunciation, other than those mentioned described above</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. This child needs help with his/her speech (e.g. speech therapy)</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td>(Circle)</td>
<td></td>
</tr>
</tbody>
</table>
Study IV, untrained listeners’ assessment form  Appendix 6

Speech assessment

Circle the appropriate number

**Speaks through the nose/ has a blocked nose**

- Normal 0
- Slight or mild but intermittent 1
- Mild but consistent 2
- Moderate 3
- Severe 4

**It sounds like puffs of air coming through the nose**

- None 0
- Once or twice 1
- Some times/several times 2
- Often 3
- Consistently or almost consistently 4

**Global evaluation**

This child needs help with his/her speech

Yes  No  (Circle)

*Intermittent* = occurring every now and then