The Nasality Severity Index 2.0

Adaptation and application of a new multiparametric approach to hypernasality

Kim Bettens

2016

Prof. dr. Kristiane Van Lierde
Prof. dr. Floris Wuyts
The Nasality Severity Index 2.0:
Adaptation and application of a new multiparametric approach to hypernasality

Kim Bettens

Supervisor: Prof. dr. Kristiane Van Lierde
Co-supervisor: Prof. dr. Floris Wuyts

Thesis submitted to fulfill the requirements for the degree of Doctor in Health Sciences

2016
Examining board

Prof. Dr. J. Van De Voorde (Chair)
Prof. Dr. P. Corthals
Prof. Dr. E. De Leenheer
Prof. Dr. G. De Pauw
Prof. Dr. E. Gerrits
Prof. Dr. E. Manders
Prof. Dr. P. Van Cauwenberghe

Cover design
Ellen Creve | ellen@bgsh.be

Printing
University Press | Leegstraat 15 | 9060 Zelzate | Belgium
Tel: +32 (0)9 342 72 25 | info@universitypress.be
www.universitypress.be

©Kim Bettens, Ghent, Belgium
All rights reserved. No part of this work may be reproduced in any form or by any means, electronically, mechanically, by print, or otherwise, without prior written permission by the author.

Ghent University
Department of Speech, Language and Hearing Sciences
De Pintelaan 185, 2P1
9000 Ghent | Belgium
Tel: +32 (0)9 332 94 26 | +32 (0)494 40 76 76
Kim.Bettens@UGent.be | kimbettens@hotmail.com
“When people of similar frequencies come together, output is not a simple sum of individual work, but exponential. In science we term this phenomenon as resonance. Output at this stage is beyond any logical limit.”

Adapted from Ravindra Shukla, A Maverick Heart: Between Love and Life, 2012
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF PUBLICATIONS</td>
<td>7</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>8</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>10</td>
</tr>
<tr>
<td>SAMENVATTING</td>
<td>12</td>
</tr>
<tr>
<td><strong>PART 1: GENERAL INTRODUCTION</strong></td>
<td>15</td>
</tr>
<tr>
<td>CHAPTER 1: Instrumental and perceptual assessment of the velopharyngeal function and resonance</td>
<td>17</td>
</tr>
<tr>
<td>CHAPTER 2: Research objectives</td>
<td>49</td>
</tr>
<tr>
<td><strong>PART 2: PUBLICATIONS</strong></td>
<td>53</td>
</tr>
<tr>
<td>CHAPTER 3: Effects of age and gender in normal-speaking children on the Nasality Severity Index: an objective multiparametric approach to hypernasality</td>
<td>55</td>
</tr>
<tr>
<td>CHAPTER 4: The Nasality Severity Index 2.0: Revision of an objective multiparametric approach to hypernasality</td>
<td>69</td>
</tr>
<tr>
<td>CHAPTER 5: Influence of gender and age on the Nasality Severity Index 2.0 in Dutch-speaking Flemish children and adults</td>
<td>91</td>
</tr>
<tr>
<td>CHAPTER 6: Short-term and long-term test-retest reliability of the Nasality Severity Index 2.0</td>
<td>107</td>
</tr>
<tr>
<td>CHAPTER 7: The relationship between the Nasality Severity Index 2.0 and perceptual judgments of hypernasality</td>
<td>125</td>
</tr>
<tr>
<td>CHAPTER 8: Short-term effect of short, intensive speech therapy on articulation and resonance in Ugandan patients with cleft (lip and) palate</td>
<td>151</td>
</tr>
<tr>
<td><strong>PART 3: GENERAL DISCUSSION</strong></td>
<td>173</td>
</tr>
<tr>
<td>CHAPTER 9: General discussion and conclusion</td>
<td>175</td>
</tr>
<tr>
<td>DANKWOORD</td>
<td>191</td>
</tr>
</tbody>
</table>
This doctoral thesis is based on the articles published or submitted in the following international peer-reviewed journals:


<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% CI</td>
<td>95% confidence interval</td>
</tr>
<tr>
<td>95% PI</td>
<td>95% prediction interval</td>
</tr>
<tr>
<td>ANA</td>
<td>audible nasal airflow</td>
</tr>
<tr>
<td>ANCOVA</td>
<td>analysis of covariance</td>
</tr>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>Arms</td>
<td>root-mean-square amplitude</td>
</tr>
<tr>
<td>BCLP</td>
<td>bilateral cleft lip and palate</td>
</tr>
<tr>
<td>CAPS-A</td>
<td>Cleft Audit Protocol for Speech – Augmented</td>
</tr>
<tr>
<td>CLP</td>
<td>cleft lip and palate</td>
</tr>
<tr>
<td>CP</td>
<td>cleft palate</td>
</tr>
<tr>
<td>CP±L</td>
<td>cleft palate with or without cleft lip</td>
</tr>
<tr>
<td>CoRSU</td>
<td>Comprehensive Rehabilitation Services in Uganda</td>
</tr>
<tr>
<td>CPAP</td>
<td>continuous positive airway pressure</td>
</tr>
<tr>
<td>CT</td>
<td>computed tomography</td>
</tr>
<tr>
<td>DME</td>
<td>direct magnitude estimation</td>
</tr>
<tr>
<td>DSI</td>
<td>Dysphonia Severity Index</td>
</tr>
<tr>
<td>EAI</td>
<td>equal appearing interval</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transformation</td>
</tr>
<tr>
<td>HFP</td>
<td>high frequency power</td>
</tr>
<tr>
<td>HONC index</td>
<td>Horii’s Oral Nasal Coupling index</td>
</tr>
<tr>
<td>ICC</td>
<td>intraclass correlation coefficient</td>
</tr>
<tr>
<td>IPA</td>
<td>international phonetic alphabet</td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board</td>
</tr>
<tr>
<td>LFP</td>
<td>low frequency power</td>
</tr>
<tr>
<td>LTAS</td>
<td>long-term average spectrum</td>
</tr>
<tr>
<td>MDD</td>
<td>minimal detectable difference</td>
</tr>
<tr>
<td>MDT</td>
<td>maximum duration time</td>
</tr>
<tr>
<td>MRI</td>
<td>magnetic resonance imaging</td>
</tr>
<tr>
<td>n.a.</td>
<td>not applicable</td>
</tr>
<tr>
<td>NORAM</td>
<td>Nasality Oral Ratio Meter</td>
</tr>
<tr>
<td>NSI</td>
<td>Nasality Severity Index</td>
</tr>
<tr>
<td>PAT-3</td>
<td>Photo Articulation Test – third edition</td>
</tr>
<tr>
<td>PTA</td>
<td>pure threshold audiometry</td>
</tr>
<tr>
<td>ROC</td>
<td>receiver operating characteristic</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation of the mean</td>
</tr>
<tr>
<td>SEM</td>
<td>standard error of measurement</td>
</tr>
<tr>
<td>SLP</td>
<td>speech-language pathologist</td>
</tr>
<tr>
<td>SNAP</td>
<td>Simplified Nasometric Assessment Procedures</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SNAP-R</td>
<td>Simplified Nasometric Assessment Procedures – Revised</td>
</tr>
<tr>
<td>SNORS</td>
<td>Super Nasal Oral Ratiometery System</td>
</tr>
<tr>
<td>SNR</td>
<td>signal-to-noise ratio</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for Social Sciences</td>
</tr>
<tr>
<td>t.n.p.</td>
<td>testing not possible</td>
</tr>
<tr>
<td>TONAR</td>
<td>The Oral Nasal Acoustic Ratio</td>
</tr>
<tr>
<td>UCLP</td>
<td>unilateral cleft lip and palate</td>
</tr>
<tr>
<td>VAS</td>
<td>visual analogue scale</td>
</tr>
<tr>
<td>VIF</td>
<td>variance inflation factor</td>
</tr>
<tr>
<td>VLHR</td>
<td>voice low tone to high tone ratio</td>
</tr>
<tr>
<td>VPI</td>
<td>velopharyngeal insufficiency</td>
</tr>
<tr>
<td>VRPlc</td>
<td>Voice Range Profile Index for Children</td>
</tr>
</tbody>
</table>
SUMMARY

For decades, researchers have been searching for the most ideal assessment technique in order to diagnose resonance disorders and to decide on the most appropriate treatment. Currently, the presence and degree of resonance disorders is determined by a combination of perceptual judgments and indirect assessment techniques. As speech perception is fundamentally perceptual in nature, perceptual assessments have been traditionally applied to evaluate speech disorders. However, several variables can influence listeners' perception of resonance which may limit the reliability and validity of perceptual judgments. Consequently, numerous indirect assessment techniques were developed to complement and objectify perceptual assessments. Nevertheless, no indirect technique can yet closely reflect the capabilities of the human ear.

A possible solution to sidestep the limitations of single indirect instrumental assessment techniques is the combination of different variables into a multiparametric index. Following this, Van Lierde et al. (2007) took a first step in creating an instrumental and multiparametric protocol to assess resonance disorders by constructing a ‘Nasality Severity Index’ (NSI). The initial aim of this doctoral thesis was to explore the application of this NSI as a new, multiparametric approach to determine hypernasality in daily clinical practice. To verify the possible influence of personal and environmental variables on the original NSI, the data of 74 Dutch-speaking Flemish children without resonance disorders (mean age (SD): boys 8.3y (2.0), girls 8.4y (2.2); range 4-12y) were analyzed. Based on these data, an influence of age and environmental variables was found, resulting in a large spread of NSI values, even in children without resonance disorders.

Consequently, an adaptation of the original NSI with inclusion of new assessment techniques was aimed, in which the influence of these above-mentioned variables was taken into account. Therefore, additional acoustic techniques to determine hypernasality were explored and included in a statistical analysis to derive a second version of the NSI, namely the Nasality Severity Index version 2.0 (NSI 2.0). Based on the optimal statistical discrimination of 35 children with perceived hypernasality and a control group of 50 children without resonance disorders, a weighted linear combination of three acoustic parameters was established. More specifically, the nasalance value of the vowel /u/ and an oral text passage obtained by the Nasometer and the voice low tone to high tone ratio (VLHR) of a sustained vowel /i/ with a cutoff frequency of 4.47*F0Hz were included. The formula of the adapted NSI yields NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%)) – (0.260 x nasalance oral text (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)). With a sensitivity of 92% and a specificity of 100%, using a cutoff score of zero, the NSI 2.0 distinctively discriminates children with hypernasality from children with normal resonance, in which a score below zero indicates the presence of hypernasality.
To implement this new index in daily clinical practice, normative values derived from children and adults without resonance disorders, short-term and long-term test-retest reliability and the relationship between perceptual judgments of hypernasality and the NSI 2.0 and its components were successively determined. Regarding the reference values for the NSI 2.0, no statistically significant influence of gender and age was detected on the NSI 2.0 and its parameters in children. However, significantly lower NSI 2.0 scores were observed in women compared to men, without an effect of age. When the data of children and adults were compared, a significant interaction between gender and age was found for the NSI 2.0 scores, in which adult men showed higher NSI 2.0 scores compared to adult women and children. Based on these study outcomes, separate reference values for the NSI 2.0 and its parameters were established for children, adult men and adult women. With an intraclass correlation coefficient (ICC) of 0.77 in children and 0.84 in adults, NSI 2.0 scores of consecutive measurements are reliable, in which a difference of 2.68 in children and 2.82 in adults can be considered as a genuine change. Additionally, a significant correlation was withheld between the perceptual judgment of hypernasality and the NSI 2.0 scores, in which a more negative NSI 2.0 score indicates the presence of more severe hypernasality. Finally, the NSI 2.0 was applied to objectify the short-term effectiveness of short, intensive speech therapy on the resonance of patients with a history of cleft palate in addition to perceptual assessments.

Considering that the NSI 2.0 can only provide information about hypernasality and the influence of audible nasal airflow on the NSI 2.0 scores is yet unclear, the inclusion of perceptual judgments in the assessment of resonance disorders remains necessary. As both assessment procedures are complementary, they can restrain each other’s limitations and may stimulate critical thinking, especially when contradictory results are observed. In the future, additional instrumental correlates of hypernasality based on connected speech could be further explored. In conclusion, the multiparametric NSI 2.0 forms a new, more powerful approach in the assessment of and treatment planning for individuals presenting with hypernasality.

References

SAMENVATTING

Onderzoekers zijn reeds lange tijd op zoek naar de ideale onderzoekstechniek om resonantiestoornissen te diagnosticeren en de meest geschikte behandeling te selecteren. Aangezien spraakwaarneming fundamenteel perceptueel van aard is, worden spraakstoornissen, en meer specifiek resonantiestoornissen, van oudsher geëvalueerd op basis van perceptuele beoordelingen. Verschillende variabelen kunnen de spraakperceptie van de luisteraar echter beïnvloeden waardoor de betrouwbaarheid en validiteit van deze onderzoekstechniek in vraag gesteld kunnen worden. Bijgevolg werden verschillende indirecte onderzoekstechnieken ontwikkeld om de perceptuele beoordeling te ondersteunen en te objectiveren. Ondanks deze inspanningen weerspiegelt op dit moment nog geen enkele indirecte onderzoekstechniek de mogelijkheden van het menselijke oor.

Een mogelijke oplossing om de beperkingen van enkelvoudige, indirecte instrumentele technieken te omzeilen is het combineren van verschillende technieken tot een multiparametrische index. Op basis van dit principe zetten Van Lierde et al. (2007) de eerste stap in de ontwikkeling van een instrumenteel en multiparametrisch protocol om resonantiestoornissen te onderzoeken door het ontwerpen van een ‘Nasality Severity Index’ (NSI). Het initiële doel van dit doctoraatsproefschrift was het verkennen van de mogelijke toepassingen van deze NSI als nieuwe, multiparametrische benadering voor het bepalen van hypernasaliteit in de dagelijkse klinische praktijk. Om de mogelijke invloed van persoonlijke en omgevingsfactoren op de oorspronkelijke NSI na te gaan, werden de data van 74 Vlaamssprekende kinderen zonder resonantiestoornissen (gemiddelde leeftijd (SD): jongens 8.3j (2.0), meisjes 8.4j (2.2); range 4-12j) onderzocht. Op basis van deze data werd een invloed van leeftijd en omgevingsfactoren weerhouden wat resulteerde in een grote spreiding van de NSI-waarden, zelfs bij kinderen zonder resonantiestoornissen.

Als gevolg hiervan werd een aanpassing van de oorspronkelijke NSI beoogd met de inclusie van nieuwe onderzoekstechnieken, rekening houdend met de invloed van persoonlijke en omgevingsfactoren. Bijgevolg werden bijkomende akoestische technieken om hypernasaliteit te detecteren onderzocht en geïncludeerd in een statistische analyse om vervolgens een tweede versie van de NSI af te leiden, namelijk de Nasality Severity Index versie 2.0 (NSI 2.0). Op basis van de optimale statistische discriminatie tussen 35 kinderen met hypernasaliteit en een controlegroep van 50 kinderen zonder resonantiestoornissen werd een gewogen lineaire combinatie van drie akoestische parameters samengesteld. Meer specifiek werden de nasaliteitswaarden van de klinker /u/ en van een orale tekst, bepaald door een Nasometer, en de voice low tone to high tone ratio (VLHR) van een aangehouden klinker /i/ op basis van een cutoff-frequentie van 4.47*F0Hz weerhouden als parameters van de aangepaste NSI. De formule van de aangepaste NSI is NSI 2.0 = 13.20 – (0.0824 x nasaliteitswaarde /u/ (%)) – (0.260 x nasaliteitswaarde orale tekst (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)). Wanneer een cutoffscore van nul gehanteerd wordt, identificeert de NSI
2.0 92% van de kinderen met hypernasaliteit en 100% van de kinderen met een normale resonantie waarbij een score onder nul overeenkomt met de aanwezigheid van hypernasaliteit.

Om deze nieuwe index toe te kunnen passen in de dagelijkse klinische praktijk werden achtereenvolgens normatieve waarden bepaald voor kinderen en volwassenen zonder resonantiestoornissen, werd de test-hertest betrouwbaarheid op korte en lange termijn onderzocht en werd de relatie tussen de perceptuele beoordeling van hypernasaliteit en de NSI 2.0 en zijn componenten bepaald. Op basis van de NSI 2.0-scores gemeten bij kinderen zonder resonantiestoornissen werd geen invloed van leeftijd en geslacht gevonden. Volwassen mannen veroorzaarden echter hogere NSI 2.0-scores in vergelijking met volwassen vrouwen, waarbij geen leeftijdseffect werd weerhouden. Bij het vergelijken van de data van kinderen en volwassenen werd een significant interactie-effect tussen geslacht en leeftijd gevonden waarbij volwassenen mannen hogere NSI 2.0 scores vertoonden in vergelijking met volwassen vrouwen en kinderen. Deze resultaten leidden tot het opstellen van afzonderlijke normwaarden voor de NSI 2.0 bij kinderen, volwassen mannen en volwassen vrouwen. Op basis van een intraclass correlatiecoëfficiënt (ICC) van 0.77 voor kinderen en 0.84 voor volwassenen kunnen opeenvolgende NSI 2.0-scores betrouwbaar vergeleken worden waarbij een verschil van 2.68 bij kinderen en 2.82 bij volwassenen als een welkome verschil beschouwd kan worden. Daarenboven werd een significante correlatie gevonden tussen de perceptuele beoordeling van hypernasaliteit en de NSI 2.0-scores waarbij een meer negatieve NSI 2.0-score overeenkomt met de aanwezigheid van meer hypernasaliteit. Ten slotte werd de NSI 2.0 toegepast om het kortetermijneffect van korte, intensieve logopedische therapie op de resonantie van patiënten geboren met een palatoschisis te objectiveren als aanvulling op de perceptuele beoordeling.

Aangezien de NSI 2.0 enkel informatie biedt over de mate van hypernasaliteit en de invloed van hoorbare nasale luchtstroom op de NSI 2.0-scores nog niet duidelijk is, blijft de inclusie van perceptuele beoordelingen bij het onderzoeken van resonantiestoornissen belangrijk. Instrumentele en perceptuele onderzoeksmethoden zijn echter complementair waarbij ze elkaars beperkingen kunnen opheffen en het kritisch denken van de clinicus kunnen bevorderen, vooral wanneer tegenstrijdige resultaten gevonden worden. In de toekomst kunnen daarenboven bijkomende akoestische correlaten van hypernasaliteit op basis van lopende spraak onderzocht worden. Besluitend kunnen we stellen dat de multiparametrische NSI 2.0 een nieuwe, meer solide techniek vormt in de diagnostiek van en therapieplanning voor personen met hypernasaliteit.

Referenties
PART 1

GENERAL INTRODUCTION
In physics, resonance may simply be defined as the oscillating response of a body or air-filled cavity to an impulse (Peterson-Falzone et al., 2001a; Wood, 1971). During speech production, resonance results from the transmission of the vocal fold vibrations from the larynx to the pharynx and oral and/or nasal cavity. During this transmission, the vibrating frequency of the vocal folds, i.e. the fundamental frequency, and its harmonics are selectively enhanced or attenuated as a function of the shape and the size of these cavities determined by personal characteristics and the movements of the speech articulators (i.e. tongue, lips, mandible, velopharyngeal mechanism) (Kummer, 2011a; Titze, 2000). Consequently, unique spectral qualities are assigned to the sound resulting in acoustically distinct speech sounds (Kummer, 2011a).

In Dutch, transmission of the vibrations through the nasal cavity will normally result in the production of one of the nasal phonemes /m/, /n/, /ŋ/ or /ɲ/. To achieve the production of all other, voiced oral phonemes, the vibrations have to be directed through the oral cavity which requires the separation of the nasal cavity from the other resonating areas. For the production of all unvoiced oral phonemes, airflow without vibrations needs to be directed from the lungs to the oral cavity which also requires a separation between the oral and nasal cavity. To accomplish this separation, the velopharyngeal valve, including the velum, the lateral pharyngeal walls, and the posterior pharyngeal wall (Kummer, 2014; Perry, 2011), has to close properly. To establish this adequate closure, three conditions have to be fulfilled: (1) the structures of the velopharyngeal mechanism have to be normal (i.e. normal anatomy), (2) these structures have to move properly (i.e. normal neurophysiology) and (3) articulation learning has to be normal (Kummer, 2014).

If (one of) these condition(s) is (are) not fulfilled, resonance and airflow disorders and, as a consequence, articulation errors and voice problems may arise (Kummer, 2008; Peterson-Falzone et al., 2001a). Those aspects play an important role in the speech intelligibility of an individual (De Bodt et al., 2002). Speech intelligibility is defined as the degree to which the speaker’s message can be understood by the listener (Henningsson et al., 2008) and is often referred to as the most important outcome measure of speech intervention, whether this includes surgery or speech therapy (Sell, 2005). Therefore, when resonance problems emerge, a multidisciplinary assessment of all these aspects, including the velopharyngeal...
structures and their function, is necessary to result in a correct diagnosis, the establishment of an individualized treatment plan and the evaluation of the applied interventions.

In this introduction, the assessment of the velopharyngeal structures and function as well as the assessment of resonance disorders will be described and discussed. Consequently, the research objectives of the current doctoral thesis will be listed in chapter two.

1. Velopharyngeal dysfunction

During speech production, the closure of the velopharyngeal mechanism is necessary to produce oral speech sounds by retraction and elevation of the velum in combination with lateral and posterior pharyngeal wall movements (Perry, 2011). Incomplete or inconsistent closure of the velopharyngeal valve can result in resonance disorders, such as hypernasality, and airflow disorders, such as audible nasal emission (see appendix A for an overview). As a reaction on this, compensatory mechanisms can arise (Kummer, 2011a), which can consist of compensatory articulation, like glottal stops and pharyngeal fricatives, and compensatory movements, for example the appearance of a nasal or facial grimace (Kummer, 2011a).

As treatment options are strongly related to the cause of velopharyngeal dysfunction, different terminology is used to assign velopharyngeal dysfunction due to different origins. Velopharyngeal insufficiency points to an abnormality of the velopharyngeal anatomy and can be due to a history of (submucous) cleft palate, a palatopharyngeal disproportion due to for example incautious adenotomy, maxillary advancement or the removal of (non)malignant growths on the palate or pharynx, or a mechanical obstruction to achieve velopharyngeal closure due to hypertrophic tonsils or an irregular shaped adenoid pad (Kummer et al., 2015; Peterson-Falzone et al., 2001b; Trost-Cardamone, 1989). Velopharyngeal incompetence indicates a deviation of the neurophysiological activation of the velopharyngeal structures. Causes of velopharyngeal incompetence include hypotonia, dysarthria, apraxia of speech, cranial nerve defects and stress incompetence (Kummer et al., 2015; Peterson-Falzone et al., 2001b; Trost-Cardamone, 1989). A last category of velopharyngeal dysfunction includes velopharyngeal mislearning which refers to an abnormal articulation placement during speech production (Kummer et al., 2015; Trost-Cardamone, 1989). This abnormal articulation placement can result from persisting compensatory speech productions after correction of the velopharyngeal structures and/or function. Another cause of velopharyngeal mislearning is the presence of hearing loss or deafness. As appropriate auditory feedback is missing during speech development, which is necessary to learn to discriminate between oral and nasal sound productions, inappropriate opening and closure of the velopharyngeal mechanism can arise resulting in resonance disorders (Baudonck et al., 2015; Kummer et al., 2015). However, some authors attribute the impression of hypernasality in deaf speech to the presence of slow speaking rate, articulation errors, unregulated voice pitch and loudness variations (Fletcher & Higgins, 1980; Fletcher et al., 1999).
When resonance and articulation errors emerge as a result of abnormal velopharyngeal structures, speech therapy alone cannot solve these speech problems as it cannot change the involved structures. Therefore, surgery to correct the velopharyngeal structures is necessary in case of velopharyngeal insufficiency. If speech problems remain after surgical correction or are not related to structural deviations, which is the case in velopharyngeal incompetence or mislearning, speech therapy can be applied. Consequently, detecting the cause of resonance and articulation errors, as well as the follow-up of speech after surgery, is essential in these patients (Kummer, 2011c).

To select the most effective treatment plan for these problems, several instruments have been developed during the last decades to subjectively or quantitatively investigate velopharyngeal function and resonance. First of all, the speech-language pathologist can perceptually assess resonance and articulation based on a speech sample consisting of a standardized articulation test, the repetition of syllables or sentences and automatic or spontaneous speech (Kummer, 2011b). In addition, perceptual tests can be applied to clarify this evaluation procedure, such as the Bzoch tests (Bzoch, 1989), used to assess hypernasality, nasal emission or hyponasality, and the Gutzmann a/i test (Gutzmann, 1913), used to evaluate hypernasality. However, several variables can influence listeners’ perception of speech which may limit the reliability and validity of perceptual judgments (Kreiman et al., 1993). Moreover, listener judgments between clinical centers are often difficult to compare because of different speech samples and various evaluation protocols (Henningsson et al., 2008). Despite these limitations, perceptual assessments still remain the gold standard in the evaluation of resonance (Henningsson et al., 2008), because no existing instrumental assessment technique can yet transcend the capabilities of the human ear. To restrict the limitations of listener evaluations, several instrumental measurements are available to supplement perceptual assessments. These instrumental assessments can be divided in two groups: direct and indirect techniques (Table 1.1). Direct techniques directly visualize the velopharyngeal closing mechanism, whereas indirect techniques provide information from which the velopharyngeal activity can be inferred.

This chapter discusses perceptual and instrumental assessment techniques presently available. Because the assessment of resonance disorders is a comprehensive topic with multiple contradictions resulting in different assessment methods, clinical application of these techniques and their advantages and limitations are not always straightforward. Therefore, the current chapter presents a structured and critical overview for clinicians who make decisions about interventions in patients with resonance disorders. Additionally, possibilities to reduce the mentioned limitations of available techniques, considering a multidimensional approach to the nasality assessment, will be discussed.
Table 1.1 Summary of direct and indirect instrumental assessment techniques discussed in this chapter.

<table>
<thead>
<tr>
<th>Direct assessment techniques</th>
<th>Indirect assessment techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasopharyngoscopy</td>
<td>Acoustic measurements</td>
</tr>
<tr>
<td>Multiview videofluoroscopy</td>
<td>- Accelerometric techniques</td>
</tr>
<tr>
<td>(Dynamic) magnetic resonance imaging</td>
<td>- Nasometry</td>
</tr>
<tr>
<td>Lateral cephalometric radiography</td>
<td>- Spectral analysis</td>
</tr>
<tr>
<td>Computed tomography</td>
<td>Aerodynamic measurements</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>- Nasal and oral airflow</td>
</tr>
<tr>
<td></td>
<td>- Pressure-flow technique</td>
</tr>
</tbody>
</table>

2. Perceptual assessment techniques

Speech perception is fundamentally perceptual in nature. A speech disorder only exists when it is recognized by the patient and/or others in the patient’s environment. Perceptual assessments of speech disorders by clinicians are based on this principle and remain the gold standard during the diagnosis and the evaluation of speech therapy and surgical intervention. Moreover, several authors consider perceptual assessments as the standard against which instrumental measurements must be validated (Kent, 1996; Keuning et al., 2004; Kreiman et al., 1993; Moll, 1964; Vogel et al., 2009). However, several variables can influence listeners’ perception of speech which may limit the reliability and validity of perceptual judgments (Kreiman et al., 1993). More specifically, individual differences due to experience, perceptual habits and biases, which determine the listener’s internal standard, can influence perceptual judgments. Additionally, task factors such as definitions of rating scales, listeners’ familiarity with the used scale and perceptual context (i.e. the ‘listener drift’ in which listeners rate speech as more severely disturbed when a moderately impaired speech sample follows a series of mildly impaired samples) may have a potential influence. Lastly, interaction between listeners’ and task factors, such as differences in the interpretation of the rating points of the used scale, can also influence listeners’ decisions.

Furthermore, different protocols are available to judge resonance, resulting in a variation of applied speech samples, recording procedures, rating scales and terminology (Henningsson et al., 2008; John et al., 2006; Lohmander & Olsson, 2004; Lohmander et al., 2009). Regarding the speech sample, several authors (Counihan & Cullinan, 1970; Henningsson et al., 2008; John et al., 2006; Kuehn & Moller, 2000; Sell, 2005) recommend the inclusion of spontaneous speech and sentence repetition, in which the use of sentence repetition includes the possibility to control for phonetic content (Sell, 2005). However, spontaneous speech and single words are the speech samples most often mentioned for perceptual analysis of cleft palate speech (Lohmander & Olsson, 2004). In order to meet the demand for a uniform speech sample, Henningsson et al. (2008) proposed universal parameters that can be applied in several languages to compose a consistent speech sample. However, the reliability and validity of those samples are not yet confirmed (Chapman et al., 2016).
Regarding the recording and presentation of the speech samples, no consensus is reached about the use of audio or video (Sell et al., 2009). Nevertheless, high quality of both the recording as well as the presentation of the speech sample in combination with a good listening environment is necessary to ensure the reliability of the ratings (Gooch et al., 2001; Sell et al., 2009).

Next to the selection and recording of an appropriate speech sample, the rating procedures need special attention. The discussion continues about the type of rating scale that has to be used. Equal appearing interval (EAI) scales with clear description of the different grades are recommended in different perceptual assessment protocols for resonance disorders in patients with cleft palate (Henningsson et al., 2008; John et al., 2006; Sell, 2005; Sell et al., 1994, 1999; Sweeney & Sell, 2008). Moreover, EAI scaling was applied in 74% of the studies that included a perceptual assessment of cleft palate speech, as reported by Lohmander and Olsson (2004). However, recent studies suggested that resonance can be rated more reliably and validly by using ratio scales such as direct magnitude estimation (DME) or visual analogue scaling (VAS) (Baylis et al., 2011, 2015; Whitehill et al., 2002; Zraick & Liss, 2000). Brancamp et al. (2010), on the other hand, reported no statistically significant differences between ratings of nasality based on EAI and DME scales.

Finally, listeners need adequate and uniform information about the terminology to describe nasality reliably (Brunnegard & Lohmander, 2007; Kent et al., 1999; Whitehill, 2002). Therefore, inclusion of reference samples and training sessions may improve consistency as the instable internal standard of the listener is then replaced by perceptual references (Chapman et al., 2016; Kreiman et al., 1993; Sell et al., 2009).

3. Direct assessment techniques

Direct assessment techniques directly visualize the velopharyngeal closing mechanism and provide information about velopharyngeal gap size and shape. Based on this information, clinicians will identify patients who could benefit from treatment, quantify the severity of velopharyngeal dysfunction, select the most suitable treatment procedure and quantify changes after treatment (Witt et al., 2000). The ideal direct assessment technique should be “noninvasive, easily repeatable, and reproducible; it should not use ionizing radiation, and should allow completely free choice of the image planes in all three dimensions” (Beer et al., 2004, p 791). Successively, the procedure, purpose, advantages and limitations of nasopharyngoscopy, multiview videofluoroscopy, (dynamic) magnetic resonance imaging (MRI), lateral cephalometric radiographic analysis, computed tomography (CT) and ultrasound will be described and discussed.

3.1 Nasopharyngoscopy

During the nasopharyngoscopic assessment, a fiberoptic scope is passed through a nostril into the nasopharynx and is placed above the velum to obtain a birds’-eye view on the velopharynx in rest and during real-time phonation (Lam et al., 2006; Rudnick & Sie,
2008; Silver et al., 2011) to identify the movement patterns during connected speech (Osberg & Witzel, 1981; Witt et al., 2000). To determine velopharyngeal disorders, a well-defined speech sample (Karnell, 2011) and good cooperation of the patient (Beer et al., 2004; Havstam et al., 2005; Silver et al., 2011; Witt et al., 2000) are necessary.

An advantage of nasopharyngoscopy is that the patient is not exposed to ionizing radiation (Kuehn & Moller, 2000; Witt et al., 2000) which induces the possibility to use this technique during biofeedback training (Berkowitz, 2013; Van Lierde et al., 2004). Furthermore, Lam et al. (2006) found a stronger correlation between the grade of velopharyngeal insufficiency and the results of the nasopharyngoscopic assessment in comparison with multiview videofluoroscopy because nasopharyngoscopy indicates more precisely the grade of velar closing.

However, the insertion of the scope is rather invasive, which can prevent children to cooperate (Beer et al., 2004; Havstam et al., 2005; Silver et al., 2011; Witt et al., 2000). Furthermore, the birds’-eye view represents only one image plane (the axial or “en face” view) (Kane et al., 2002; Sinclair et al., 1982), which does not always provide an optimal view to the movements of the pharyngeal walls because of velar or adenoidal blockage (Henningsson & Isberg, 1991; Witt et al., 2000). Additionally, the visualization of underlying structures and muscles is not possible (Bae et al., 2011; Kane et al., 2002), and although qualitative evaluations of the velopharyngeal function can be performed, a quantitative analysis is difficult because of the two-dimensional representation of the three-dimensional anatomy (Kane et al., 2002; Lam et al., 2006; Silver et al., 2011).

3.2 Multiview videofluoroscopy

Multiview videofluoroscopy is another technique to visualize the structure, movements, closing and timing of the velopharyngeal mechanism (Havstam et al., 2005). After a suspension of colloidal barium sulphate is applied to coat the nasopharynx to increase contrast, fluoroscopic images are acquired in lateral, frontal and base (“en face”) or Towne views (Kane et al., 2002; Silver et al., 2011; Skolnick, 1970; Stringer & Witzel, 1986) during connected speech. The simultaneous view of all articulators (Shprintzen, 2013) supplies unique information about abnormal compensatory movements, such as tongue-backing and possible paradoxical movements of the velopharynx, as well as timing abnormalities (Lam et al., 2006).

By obtaining the lateral view, abnormalities of the velum, posterior pharyngeal wall and tongue as well as the presence of any Passavant ridge are easy to visualize (Stringer & Witzel, 1986; Witt et al., 1998). Furthermore, the height of velopharyngeal closure can be identified with reference to the vertebrae (Silver et al., 2011) and the size of tonsils can be determined (Witt et al., 2000). The frontal view offers the possibility to demonstrate lateral wall movements (Witt et al., 2000). However, due to overlapping structures, it is sometimes difficult to analyze the images to collect sufficient information about the sphincteric closure.
Instrumental and perceptual assessment of the velopharyngeal function and resonance

(Stringer & Witzel, 1986). Therefore, a third view, such as the basal or Towne projection is necessary. The basal view reveals the relationship between the velum and the lateral-posterior aspects of the pharyngeal wall and is analogous to the nasopharyngoscopic view (Witt et al., 2000). It also provides information about the closure pattern of the velopharyngeal mechanism which has proved its clinical usefulness for both diagnosis and post-treatment evaluation (Witt et al., 2000), but is difficult to obtain and interpret in patients with large adenoids (Skolnick et al., 1973). In this case, the Towne projection can form a solution. Stringer & Witzel (1986) also found this position to be more comfortable for children to maintain.

Although different projections are obtained during multiview videofluoroscopy, the three-dimensional anatomy of the velopharynx is reduced to a two-dimensional image which can create an overestimation of the velopharyngeal closing (Silver et al., 2011). Moreover, it can be challenging to interpret the images because of multiple shadows (Kane et al., 2002; Silver et al., 2011) and visualization of underlying muscles is not feasible (Bae et al., 2011). Even though the time to capture images is restricted, patients are exposed to ionizing radiation (Bae et al., 2011; Beer et al., 2004; Kane et al., 2002; Karnell, 2011; Silver et al., 2011). Finally, injecting the barium sulphate can raise resistance in young children (Beer et al., 2004; Silver et al., 2011), although multiview videofluoroscopy is experienced as less invasive compared to nasopharyngoscopy (Kuehn & Moller, 2000).

3.3 (Dynamic) magnetic resonance imaging

Magnetic resonance imaging (MRI) is a more recent imaging technique that offers the possibility to collect images of different plane views (Beer et al., 2004; Drissi et al., 2011; Silver et al., 2011; Vadodaria et al., 2000) with high spatial resolution and superior visualization of soft tissues (Atik et al., 2008; Bae et al., 2011; Kane et al., 2002; Ozgür et al., 2000).

The mid-sagittal view provides information about length, movement and extensibility of the velum, forward movement of the posterior pharyngeal wall during velopharyngeal closure and the presence or absence of a Passavant ridge (Vadodaria et al., 2000). The coronal view shows the width of the pharynx and the contribution of the lateral pharyngeal wall in the velopharyngeal closure (Vadodaria et al., 2000). Finally, the axial view at the level of the hard palate offers information about type and extent of velopharyngeal closure (Vadodaria et al., 2000).

One of the main advantages of MRI is that no ionizing radiation is used, so the assessment is repeatable (Drissi et al., 2011; Rowe & D'Antonio, 2005). Recently, Maturo et al. (2012) described the possibility to synchronize audio and video samples from connected speech without delay, which creates the opportunity for real-time dynamic visualization in combination with perceptual assessment. Furthermore, MRI can also be used for the early assessment of submucous cleft palate (Perry et al., 2012).
Limitations of MRI are the high costs (Atik et al., 2008; Perry et al., 2011; Silver et al., 2011) together with the difficulty to test young children. Tian et al. (2010) reported that children from the age of 5 years old can participate without anesthesia if they receive preparatory instructions. However, fear, noise and lying quiet for a long time can cause uncooperativeness (Bae et al., 2011; Beer et al., 2004; Kane et al., 2002; Ozgür et al., 2000; Silver et al., 2011; Vadodaria et al., 2000). Several authors (Beer et al., 2004; Ettema et al., 2002; Silver et al., 2011) also indicate the possible consequences of gravity on the velum because of the supine position during the assessment. However, Perry et al. (2011) and Kollara and Perry (2014) found only minimal effect of gravity on velar thickness, velar length, velar height, levator muscle length, angles of origin, and pharyngeal dimensions. At last, fixed intraoral metallic prostheses can affect the image quality negatively (Drissi et al., 2011; Ozgür et al., 2000; Vadodaria et al., 2000).

3.4 Lateral cephalometric radiographic analysis

Lateral cephalograms taken by x-rays and analyzed by specific software programs can be obtained to visualize the anatomy of the velopharyngeal structures at rest or during sustained phonation (Jakhi & Karjodkar, 1990; Stellzig-Eisenhauer, 2001; Witt et al., 2000). With minimal patient compliance, standardized information about the relation of the soft tissues of the nasopharynx to the bony landmarks of the face and the cranium can be obtained (Kuehn & Moller, 2000; Witt et al., 2000). However, the interpretation of this information can be hampered by the presence of multiple shadows and patients are exposed to ionization radiation (Witt et al., 2000). Additionally, only limited and static information about the physiology of the velopharyngeal mechanism in the midsagittal plane can be offered by this technique which also reduces the three-dimensional anatomy into a two-dimensional representation (Atik et al., 2008; Berkowitz, 2013; Stellzig-Eisenhauer, 2001; Witt et al., 2000).

3.5 Computed tomography

Computed tomography scans (CT scans) can provide information about the anatomy of the velopharyngeal system in the axial plane in rest and during sustained phonation (Beer et al., 2004; Honjo et al., 1984). More specifically, the images can determine the level of velopharyngeal closure in relation to other soft tissues (Honjo et al., 1984) and can quantify surficial and deep craniofacial structures (Suri et al., 2008). However, the limitations of this technique are comparable with those of cephalometric measurements: patients are exposed to ionizing radiation and only limited and static information about the velopharyngeal mechanism is obtained in a two-dimensional way (Atik et al., 2008; Beer et al., 2004; Honjo et al., 1984). Despite the higher radiation exposure, recent developments in 3D and 4D-CT imaging are promising as these images provide more accurate information about the 3D anatomy and movement of the velopharyngeal mechanism (Sakamoto et al., 2015).
3.6 Ultrasound

To visualize the anatomy and physiology of the lateral pharyngeal walls, ultrasound can be applied (Hawkins & Swisher, 1978). Therefore, a transducer in combination with the use of an acoustic coupling gel has to be placed against the neck, under the ear or behind the ramus of the mandible (Hawkins & Swisher, 1978).

Despite the noninvasive character of this technique without ionizing radiation (Ryan & Hawkins, 1976), ultrasound offers too little advantages in the assessment of the velopharyngeal mechanism because of the restricted visibility of the (motion of) the velum (Baken & Orlikoff, 2000; Berkowitz, 2013; Hawkins & Swisher, 1978; Ozgür et al., 2000) and too high interobserver and intersubject variability (Ryan & Hawkins, 1976).

4. Indirect assessment techniques

Indirect assessment techniques provide information from which the velopharyngeal activity and possible malfunction can be inferred (Baken & Orlikoff, 2000). In contrast to direct techniques, these assessments aim to provide quantitative information about the degree of resonance or abnormal functioning of the velopharyngeal mechanism. Two categories of evaluation techniques can be distinguished to indirectly assess resonance and nasal airflow respectively: acoustic and aerodynamic measurements.

4.1 Acoustic measurements

To determine the presence of resonance disorders, the measurement and analysis of the sounds a patient produces afford useful information. Therefore, several researchers are looking for acoustic parameters that can discriminate between resonance disorders and normal resonance. Acoustic measurements are frequently based on the relation between nasal and oral acoustic energy which can be a correlate of perceived nasality. Within this category, the use of accelerometric techniques, nasometry and spectral analysis will be discussed.

4.1.1 Accelerometry

Horii (1980) evolved the Horii’s Oral Nasal Coupling (HONC) Index in which the ratio of the nasal accelerometer signal amplitude to the laryngeal accelerometer signal amplitude is computed to reduce the influence of vocal intensity. Two accelerometers are placed on the external surface of a speaker’s nose and throat. The index is expressed as

\[
\text{HONC} = \frac{\text{Arms}(n)}{(k \times \text{Arms}(v))}
\]

where \(\text{Arms}(n)\) is the root-mean-square amplitude of the nasal accelerometer signal, \(\text{Arms}(v)\) is the root-mean-square amplitude of the vocal accelerometer signal and \(k\) is a constant which corresponds with a HONC value of one during sustained phonation of the
sound /m/ (Horii, 1980). Due to the implementation of this constant, which compensates for variations of the accelerometer signals and interindividual differences, comparisons between and in speakers are possible. The HONC index ranges from zero to one, in which zero represents a complete oral signal and one represents the signal of a sustained sound /m/, or can be expressed in dB (Horii, 1980).

The index can differentiate between normal resonance and hypernasality, as well as between oral and nasal texts (Horii & Lang, 1981; Mra et al., 1998; Sussman, 1995); it has a moderate to strong correlation with perceptual evaluation (Horii, 1980; Horii & Lang, 1981; Laczi et al., 2005) and good interobserver reliability (Mra et al., 1998; Sussman, 1995). Furthermore, the assessment can be done based on sustained vowels or running speech and is noninvasive (Horii, 1980). However, the HONC index is rarely used in clinical or research settings (Kuehn & Moller, 2000) because it is not commercially available as a preassembled package (Laczi et al., 2005).

The Nasality Oral Ratio Meter (NORAM) (Karling et al., 1985) also utilizes nasal and laryngeal accelerometers by which the duration of the nasal (tN) and oral (tL) signal is measured and the ratio is calculated with the formula

\[ n = \frac{tN}{tL \times 100} \]

where n is the percentage of nasalance. Although NORAM can be used to examine nasalance before and after therapy (Karling et al., 1993; Lohmander-Agerskov et al., 1996), the low inter- and intraobserver reliability and the impossibility to distinguish normal resonance from hypernasality (Karling et al., 1993) cause a limited clinical and research application of this technique (Baken & Orlikoff, 2000).

### 4.1.2 Nasometry

Another acoustic measurement technique that is based on the relation between nasal and oral acoustic energy is The Oral Nasal Acoustic Ratio (TONAR), originally developed by Fletcher and Bishop (1970) and marketed by KayPentax (formerly known as Kay Elemetrics Corporation, NJ, Lincoln Park) as the Nasometer. To determine the percentage of nasalance, a plate on a headset is fixed between the upper lip and the nose of the patient. Two microphones, one on the upper side and one on the underside of the plate, record the nasal and oral signal by a connected computer. The signals are then passed through a filter with a center frequency of 500Hz and a bandwidth of 300Hz. The nasal and oral signals are compared to each other and a nasalance percentage is calculated with the formula

\[ \text{nasalance\%} = \frac{\text{nasal signal}}{\text{nasal + oral signal}} \times 100 \]
The results range from 0 to 100%. To detect hypernasality, a speech sample without nasal consonants can be used. To detect hyponasality, a speech sample loaded with nasal consonants is necessary.

The Nasometer is applied in several clinical centers and constituted the subject of sundry studies (Shprintzen & Marrinan, 2009; Vijayalakshmi et al., 2007). As a quantitative measurement technique (Karnell, 2011), it has a good test-retest reliability (Watterson & Lewis, 2006), is noninvasive, convenient and easy to interpret. Therefore, the device can also be used during feedback training (Van Lierde et al., 1999). Subtle differences for language (Nichols, 1999; Okalidou et al., 2011; Seaver et al., 1991; Van Lierde et al., 2001), gender (Rochet et al., 1998; Seaver et al., 1991; Van Lierde et al., 2003), age (Brunnegard & van Doorn, 2009; Haapanen, 1991; Rochet et al., 1998; Van Lierde et al., 2003) and race (Mayo et al., 1996) were found. Although loudness (Watterson et al., 1994), speech rate (Gauster et al., 2010) and type of oral consonants used in the speech sample (Watterson et al., 1998) have no significant effect on nasalance scores, the type of vowels can influence the results (Lewis et al., 2000). Lewis et al. (2000) mention that high vowels can lead to higher nasalance results. Several authors state good correlations with perceptual judgment (Hardin et al., 1992; Sweeney & Sell, 2008; Watterson et al., 1998); however, different cutoff scores are used to determine sensitivity and specificity of the Nasometer (Brancamp et al., 2010, 22%; Dalston et al., 1991, 32%; Hardin et al., 1992, 26%; Sweeney & Sell, 2008, 35%; Watterson et al., 1998, 26%). Moreover, it is difficult to compare these different cutoff scores because of the application of different methods across studies (Watterson et al., 1998). Some authors, on the other hand, report low correlations with perceptual assessment (Karnell, 2011; Keuning et al., 2002; Nellis et al., 1992), which leads to a disagreement about the validity of the Nasometer.

4.1.3 Spectral analysis

New techniques to determine resonance disorders are introduced thanks to the current trend of digitalization approaching resonance by using digital processing techniques which are based on specific spectral characteristics of hypernasal speech. In literature, several characteristics are described, more specifically the introduction of pole-zero pairs in the region of the first formant (Maeda, 1982; Pruthi et al., 2007; Schwartz, 1968; Vijayalakshmi et al., 2007), reduction of the amplitude of the first formant (Fant, 1970; Pruthi et al., 2007; Schwartz, 1968), increase of bandwidth of the first and second formant (Fant, 1970; Pruthi et al., 2007), shifts in formant frequencies (Fant, 1970; Hawkins & Stevens, 1985; Pruthi et al., 2007; Schwartz, 1968), rise in the amplitude between the first and second formant (Chen, 1997; Kataoka et al., 1996; Kataoka et al., 2001; Lee et al., 2003; Vijayalakshmi et al., 2007; Yoshida et al., 2000) and decrease in the amplitude at or above the second formant (Kataoka et al., 1996; Kataoka et al., 2001; Lee et al., 2003; Yoshida et al., 2000). Based on these spectral characteristics, computer logarithms are composed to discriminate between normal and hypernasal speech. Examples of such digital processing
techniques are a one-third octave spectrum analysis (Kataoka et al., 1996; Kataoka et al., 2001), the linear predictive model developed by Rah et al. (2001), a group delay-based formant extraction method (Vijayalakshmi et al., 2007) and the voice low tone to high tone ratio (VLHR), developed by Lee et al. (2003). For further information about the technical detail we refer to the listed articles.

The collection of spectral characteristics is objective, noninvasive and cost-effective (Vijayalakshmi et al., 2007). The equipment only consists of a good quality microphone, a computer with an analog-to-digital converter (Vijayalakshmi et al., 2007) and free software (Maier et al., 2008). Nevertheless, the features are sometimes difficult to interpret (Baken & Orlikoff, 2000), are influenced by loudness or individual differences (Kataoka et al., 1996; Yoshida et al., 2000) and can only be extracted from vowels. Moreover, the effect of therapy cannot be verified because no degree of hypernasality is determined (Cairns et al., 1996) or the described correlations with nasalance scores are rather low (Rah et al., 2001).

4.2 Aerodynamic measurements

Aerodynamic measurements can also be performed to evaluate the function of the velopharyngeal closure mechanism. Based on the principle that insufficient closure of the velopharyngeal mechanism causes increased nasal air escape (Warren, 1967), several tests measuring airflow and air pressure have been evolved to provide more information about the velopharyngeal functioning. Successively, techniques based on nasal and oral airflow and the pressure-flow technique (Warren & DuBois, 1964) will be described and discussed.

4.2.1 Nasal and oral airflow techniques

To visualize the amount of nasal air escape, the mirror-fogging test by Glatzel (Foy, 1910) can be used. During this test, a cold mirror is held 0.5cm under the nose of the subject during phonation of vowels or consonants. According to Glatzel, the degree of condensation is represented by four concentric circles (0-4) by which 0 corresponds to no condensation and 4 to severe condensation. The modified Glatzel mirror includes more than 4 concentric circles with 1cm distance between the lines by which the degree of condensation can be calculated by the mathematical formula for an ellipse, \( S=\pi a b \) (Brescovici & Roithmann, 2008; Gertner et al., 1984).

Although the application of this technique is simple, noninvasive and inexpensive, the reliability and validity is questionable (Brescovici & Roithmann, 2008; Pochat et al., 2012). This can be due to influencing variables such as temperature, air humidity, resistance of the nasal airways and tilting errors (Brescovici & Roithmann, 2008). However, all of these studies only assessed nasal breathing without speech production.

A more complex, but still user-friendly device to determine nasal emission is the aerophonoscope (Devani et al., 1999; Rineau, 1993). Three airflow sensors, one for each nostril and one for the oral airflow, enable the simultaneous visualization of both, nasal and
oral airflow, as well as voice levels. Additionally, an understandable graphic form displays these data qualitatively which offers the possibility to use the devise during feedback training (Devani et al., 1999). A limitation of this device, however, is that the handpiece is held under the nose in front of the mouth which can influence speech movements.

Dotevall et al. (2001) examined nasal airflow dynamics during the velopharyngeal closure phase in speech by using a pneumotachograph attached to a nasal continuous positive airway pressure (CPAP) mask. They stated that the detection of nasal airflow patterns during velopharyngeal closure is a sensitive method to determine velopharyngeal functioning quantitatively so that it can accurately distinguish between perceptually abnormal and normal resonance in children with and without cleft palate (Dotevall et al., 2001; Dotevall et al., 2002). Despite the good correlation with perceptual evaluation of hypernasality, this technique cannot determine a degree of hypernasality (Dotevall et al., 2001). Additionally, the nasal mask may have an influence on the sensory feedback for speech activity (Dotevall et al., 2002).

Another possibility to measure nasal and oral airflow is the use of a warm-wire anemometer attached to a facial mask. Airflow causes a cooling of the electrical heated wire filament by which higher velocities of airflow create a major cooling effect. To maintain the fixed temperature ratio of the heated filament more current is needed. This current is shown on a display to make interpretations possible (Quigley et al., 1964). This system, however, has a limited response speed, so it is unable to detect rapid movements of the velum (Main et al., 1999). Therefore, the Super Nasal Oral Ratiometry System (SNORS), by using high speed sensors, can overcome this deficit (Main et al., 1999). A modified oxygen mask, including airflow sensors and microphones, registers the amount of nasal airflow as a percentage of total airflow which can be displayed together with a speech envelope on a computer screen (Main et al., 1999; McLean et al., 1997). Due to its noninvasive and inexpensive character, this device can be applied during feedback training (Main et al., 1999; McLean et al., 1997).

However, it is important to realize that nasal airflow during speech can be influenced by many factors such as nasal pathway resistance, velopharyngeal airway resistance, oral air pressure, amount of air release from the lungs and respiratory effort (Warren, 1967). Furthermore, the amount of nasal emission is strongly influenced by articulatory movements and tongue position (Machida, 1967; Selley et al., 1987).

4.2.2 Pressure flow technique

To expand the measurements of nasal airflow, Warren and DuBois (1964) developed the pressure-flow technique to objectively evaluate velopharyngeal function during speech production. Based on a modification of the Theoretical Hydraulic Principle, the rate of nasal airflow in combination with the differential pressure across the velopharyngeal orifice can determine the area of that orifice. The equation yields
orifice area = \frac{\text{volume rate of airflow through the orifice}}{0.65 \sqrt{\left(2 \times (\text{intraoral air pressure} - \text{nasal air pressure})\right) \text{density of air}}}

with 0.65 a correction factor to account for “unsteady, non-uniform, and rotational” characteristics of airflow during speech production (Warren & DuBois, 1964). However, Yates et al. (1990) assume that this correction factor may be significantly higher than 0.65 based on the influence of the inlet shape of the orifice and depending on the orifice geometry.

To collect the requisite data simultaneously, two flexible catheters, one within the mouth and another in the nostril, collect intraoral and nasal air pressure (mm H\textsubscript{2}O) respectively and transmitting these pressures to pressure transducers. Furthermore, airflow is measured (ml/s) by a pneumotachograph connected by plastic tubing to the patient’s other nostril (Warren & DuBois, 1964). To make this procedure more comfortable, a nasal mask can be used to collect airflow and nasal air pressure (Gauster et al., 2010). Beside information about nasal airflow rate, oral and nasal air pressure levels and velopharyngeal orifice size, the pressure-flow technique also provides information about timing characteristics associated with speech (Warren et al., 1985).

However, the needed, specialized equipment is often not available and the procedures are technically complex and require substantial cooperation (Karnell, 2011; Kuehn & Moller, 2000). Therefore, the described technique has probably more significance for research than for daily clinical practice (Karnell, 2011).

5. Discussion

Current diagnosis of resonance disorders is based on a combination of perceptual assessments, information about the anatomy and functioning of the velopharyngeal mechanism obtained by direct assessment techniques and data from indirect techniques providing additional information regarding acoustic and aerodynamic features in speech. Appendix B provides a summary of the described assessment techniques, including their advantages and limitations as discussed above.

As no instrumental measurement meets or transcends the capabilities of the human ear yet, perceptual judgments of resonance remain the gold standard. However, assuring the reliability and validity of these procedures is still an important challenge due to the significant influence of the applied protocol, procedures for data collection and analysis, and listeners’ characteristics. It is encouraging that recent research focuses on the assessment of the reliability and validity of available protocols in which researchers work together to construct a uniform protocol including reference samples and training facilities for different languages to improve the perceptual analysis of resonance (Baylis et al., 2015; Chapman et al., 2016; John et al., 2006; Lohmander et al., 2009).
Regarding the direct assessment techniques, no ideal imaging method is yet available (Beer et al., 2004; Silver et al., 2011). Currently, nasopharyngoscopy and multiview videofluoroscopy are the most convenient instruments to assess velopharyngeal dysfunction in daily clinical practice (Lam et al., 2006). Both techniques are complementary (Lam et al., 2006) and it depends on the cause and the severity of the speech disorder whether nasopharyngoscopy or multiview videofluoroscopy is preferred (Havstam et al., 2005; Lam et al., 2006; Rowe & D’Antonio, 2005). Because these assessments rely on subjective interpretation of qualitative visual information, reliability is not always acceptable. Therefore, in 1990, a multidisciplinary international workgroup introduced a standardized rating scale to report the outcomes of instrumental assessments for velopharyngeal disorders (Golding-Kushner, 1990). However, the intrarater reliability is still too low for the scale to be used for intercenter comparisons (Sie et al., 2008). Furthermore, neither nasopharyngoscopy nor multiview videofluoroscopy matches the ideal imaging method, so innovative application of existing technologies, such as magnetic resonance imaging (MRI), creates new opportunities. In the future, a combination of dynamic MRI and audio recordings will probably provide the possibility to discriminate between normal and abnormal velopharyngeal closure based on both neuromuscular and anatomic analysis (Silver et al., 2011). However, refinement and analysis of these techniques as well as normative imaging characteristics are necessary before MRI can accurately diagnose velopharyngeal dysfunction with high sensitivity and specificity (Maturo et al., 2012).

As seen in the overview of the indirect measurement techniques, elaborate information about the velopharyngeal functioning can be derived from these methods. An extensive choice of techniques is available ranging from easily applicable and available assessments to expensive equipment in combination with complex implementation. A widespread acoustic technique to assess nasalance is the Nasometer (Fletcher & Bishop, 1970; Shprintzen & Marrinan, 2009), although no consensus exists about the correlation with perceptual assessment (Keuning et al., 2002; Sweeney & Sell, 2008). Possible explanations are that hypernasality may be rated more severely when articulation errors are present (Keuning et al., 2002) or that the human ear assesses a larger speech sample and a wider range of speech characteristics compared to the Nasometer (Karnell, 2011; Keuning et al., 2002). This disagreement about the correlation between nasalance and perceptual judgments confirms the persistent need to combine quantitative, indirect measurement with perceptual assessment (Bressmann et al., 2000; Hardin et al., 1992; Keuning et al., 2002).

Additionally, different personal variables, such as nasal pathway resistance, oral air pressure, respiratory effort (Warren, 1967) and the amount of nasal emission (Machida, 1967; Selley et al., 1987), can influence the outcome of acoustic and especially aerodynamic measurements so that the usefulness of some techniques is questionable (Warren, 1967). Therefore, it is important to be aware of these influencing factors when interpreting the test results of for example the Horii’s Oral Nasal Coupling (HONC) index, the Nasality Oral Ratio
Chapter 1

Meter (NORAM) or the pressure-flow technique (Baken & Orlikoff, 2000; Karnell, 2011; Kuehn & Moller, 2000).

According to several authors (Shprintzen & Bardach, 1995; Van Lierde et al., 2007), these subjective and quantitative measurements have to be interpreted with care and can lead to contradictory results when examining the nasality of an individual. The limitations of single assessment methods can possibly be overcome by combining several types of complementary measures.

A possible solution is the simultaneous combination of direct and indirect assessment techniques during a specific speech task, for example dynamic MRI in combination with audio recordings or SNORS+ (Sharp et al., 1999). Dynamic MRI with simultaneous audio recordings can investigate the acoustic-physiologic relation between the velopharyngeal mechanism and specific speech samples (Bae et al., 2011) or can evaluate velopharyngeal closure (Maturo et al., 2012; Silver et al., 2011). The goal of this procedure is to individualize speech therapy or surgery based on the detected functional anatomic defect (Maturo et al., 2012). However, as mentioned above, this is an expensive and time-consuming procedure and the influence of gravity on the velum is not yet clear.

The SNORS+ system (Sharp et al., 1999) includes the images of videofluoroscopy, nasopharyngoscopy, waveforms of electrolaryngography, data of nasal anemometry (SNORS) and tongue-palate contact based on electropalatography. All these techniques cover the functioning and coordination of the key articulators: velum, tongue, pharynx and larynx (Sharp et al., 1999). The data are simultaneously provided on a computer screen, allowing to interpret a combination of the assessment results. This real-time display with direct visual feedback can be used in therapy or in a diagnostic setting. Following the authors (Sharp et al., 1999), it is a clinical, user-friendly device. However, the afore-mentioned limitations of the included direct and aerodynamic assessment techniques remain: ionization radiation, two-dimensional representation of the velopharyngeal mechanism, influencing variables such as nasal pathway resistance, velopharyngeal airway resistance and nasal emission, discomfort and influence of sensory feedback due to the nasal mask (anemometry) and the artificial palate (electropalatography). Furthermore, the artificial palate can have a negative impact on speech and saliva production (McLeod & Searl, 2006).

Another multi-dimensional approach to assess resonance by the combination of only indirect techniques was presented by our research group, more specifically by Van Lierde et al. (2007). They comprised acoustic and aerodynamic characteristics obtained from an individual in one outcome measure to provide a more comprehensive descriptive of the resonance disorders and nasal airflow, more specifically the Nasality Severity Index (NSI). The NSI was based on the optimal statistical discrimination between healthy children and those with hypernasality due to a history of cleft palate, and constructed in a stepwise statistical approach, with sensitivity and specificity as the serving criteria. Based on the data set, the NSI has a sensitivity of 88% and a specificity of 95%. The five parameters used in the
NSI calculation are the nasalance scores of respectively the vowel /a/, an oral and an oronasal passage. Next, the maximum duration time (MDT) of /s/ was used as well as the mirror-fogging test by Glatzel (Foy, 1910) during the production of /a/. All digital values were determined by the Nasometer (Fletcher & Bishop, 1970, Kay Elemetrics Corporation, NJ, Lincoln Park). The equation yields \( \text{NSI} = -60.69 - (3.24 \times \text{nasalance oral text} \%) - (13.39 \times \text{Glatzel value /a/}) + (0.244 \times \text{MDT /s/ (s)}) - (0.558 \times \text{nasalance /a/ (\%)}) + (3.38 \times \text{nasalance oronasal text} \%). \) The more negative the NSI value, the higher the degree of hypernasality.

Such an index allows to implement measurements complementary to nasalance values that should be considered in the judgment of nasality as recommended by several authors (Dalston et al., 1991; Seaver et al., 1991; van Doorn & Purcell, 1998). As no specific nasal stimuli are included in the NSI and logistic regression analysis was based on the optimal discrimination of patients with hypernasality and normal resonance, the NSI cannot provide any information about the degree of hyponasality. However, hypernasality affects speech intelligibility and acceptability more than hyponasality does and therefore is clinically more relevant (Shprintzen et al., 1979). Although the clinical usefulness of the NSI has been shown by some case reports by Van Lierde et al. (2007), further research is required to confirm the possibility of the NSI to discriminate between hypernasality and normal resonance in a larger group of participants with a wider age range and to determine the relationship between the NSI scores and perceptual judgments. To advance the validity and reliability of the index, adaptation of the included parameters may be required following the aforementioned limitations considering the mirror-fogging test by Glatzel (Foy, 1910). Furthermore, it will be necessary to determine normative values before the NSI can be implemented in daily clinical practice. Because this index is only based on indirect assessment techniques, perceptual judgments and direct assessment techniques will be indispensable in treatment planning to provide information about the anatomy and functioning of the velopharyngeal mechanism.
Appendix A - Resonance and airflow disorders.

During speech production, airflow from the lungs and vibrations produced by the vocal folds are transmitted through the pharynx and oral and/or nasal cavity. The structures of these cavities together with the sound energy balance in these areas determine whether the quality of resonance is perceived as normal or deviant (Kummer, 2011a). Any disruption or obstruction of the sound transmission in the vocal tract will result in the perception of various types of resonance disorders which are described below.

Hypernasality is a deviation of resonance due to excessive resonance into the nasal tract during the production of voiced oral sounds (Kummer, 2008; Peterson-Falzone et al., 2001a). Due to the coupling of the nasal cavity to the resonating area, the acoustic filter changes, resulting in a redistribution of the sound energy which is perceived as a muffled, unclear sound with low intensity. As vowels are produced by altering the oral resonance and are relatively long in duration, hypernasality is most perceptible on these sounds. Moreover, it is more observed on vowels produced with a high tongue position, such as the vowel /i/, as this tongue position reduces the oral resonance space and induces increased sound pressure, causing an increase of sound transmission through the velum. Voiced oral consonants on the other hand may rather be perceived as their nasal counterparts (Kummer, 2011a).

Hyponasality is the reduction of normal nasal resonance during the production of nasal consonants and their adjacent vowels due to an obstruction in the nasopharynx or the nasal cavity. This obstruction can be the result of a swelling of the nasal mucosa due to a cold or allergic rhinitis, adenoid or tonsil hypertrophy, a deviated nasal septum, a velopharyngeal flap or apraxia of speech (Kummer, 2011a; Sweeney, 2011).

Cul-de-sac resonance occurs when a closed resonating cavity is coupled to the oral resonating cavity or when the sound is partially blocked from leaving that oral cavity. Depending on which cavity exit point is blocked, a distinction is made between oral, nasal or pharyngeal cul-de-sac resonance. As all sounds are partly absorbed by the soft tissues, this results in a muffled speech with low intensity (Kummer, 2008; 2011a).

Mixed resonance appears when different types of the above described resonance disorders occur at different times in connected speech. For example when an obstruction of the nasal cavity arises due to a common cold in combination with velopharyngeal insufficiency, or in patients with apraxia of speech who have difficulties with opening and closing the velopharyngeal valve at an appropriate time (Kummer, 2011a; Peterson-Falzone et al., 2001a; Sweeney, 2011).

In case of a velopharyngeal dysfunction, the airflow, whether or not vibrating, escapes through the nose. This escape can be audible or inaudible. When the release is audible, additional airflow disorders can be perceived. These disorders are discussed below.
Nasal emission results from air that escapes through the nose due to an incomplete closure of the velopharyngeal valve during the production of oral sounds or the presence of an oronasal fistula (Peterson-Falzone et al., 2001a). Especially during the production of pressure consonants (i.e. plosives, fricatives and affricates), nasal emission may be most present as the production of these sounds requires the build-up of intraoral air pressure. When hypernasality is heard, this is always accompanied by nasal emission, as air is released through the nose. However, this emission is not always audible. In contrast, nasal emission can be present without hypernasality (Kummer, 2011a).

Nasal turbulence is perceived when the air is pushed through a small velopharyngeal opening and is released on the nasal side of this opening resulting in extraturbulent noise or causing bubbling of secretions (Kummer, 2011a; Peterson-Falzone et al., 2001a). Such as nasal emission, nasal turbulence is mostly heard on pressure consonants.
## Appendix B – Summary of the perceptual and instrumental assessment techniques including their advantages and limitations.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Procedure</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perceptual assessment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceptual assessment</td>
<td>Judgment of different aspects of resonance and airflow disorders based on a live or recorded speech sample using a specific rating scale depending on the applied protocol</td>
<td>- Easy to use&lt;br&gt;- Inexpensive, rapid, noninvasive&lt;br&gt;- Meets the perceptual nature of resonance</td>
<td>- Varying reliability and validity due to influencing variables (personal and task factors)&lt;br&gt;- Several protocols including different speech samples, terminology, recording and rating procedures, etc.&lt;br&gt;- Training necessary</td>
</tr>
<tr>
<td><strong>Direct assessment techniques</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasopharyngoscopy</td>
<td>A fiberoptic scope is placed through a nostril into the nasopharynx and above the velum</td>
<td>- No ionizing radiation&lt;br&gt;- Identification of velopharyngeal function during speech, especially the grade of velar closure&lt;br&gt;- Can be used during biofeedback training</td>
<td>- Only axial view&lt;br&gt;- Invasive for children&lt;br&gt;- Two-dimensional representation of the three-dimensional anatomy</td>
</tr>
<tr>
<td>Multiview videofluoroscopy</td>
<td>After high-density barium is applied to coat the nasopharynx, fluoroscopic images are taken in lateral, frontal and base or Towne’s views</td>
<td>- Identification of velopharyngeal function during speech&lt;br&gt;- Different views providing unique information about abnormal compensatory movements, timing abnormalities, height of velopharyngeal closure (lateral view), lateral wall movements (frontal view), closure pattern (base or Towne view)&lt;br&gt;- Simultaneous view of all articulators</td>
<td>- Ionization radiation&lt;br&gt;- Interpretation can be difficult because of multiple shadows&lt;br&gt;- No visualization of underlying muscles&lt;br&gt;- Two-dimensional representation of the three-dimensional anatomy</td>
</tr>
</tbody>
</table>
### Appendix B – (continued)

<table>
<thead>
<tr>
<th>Technique</th>
<th>Procedure</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct assessment techniques (continued)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>(Dynamic)</strong> magnetic resonance imaging</td>
<td>Magnetic resonance images in different plane views using an MRI scanner (in combination with a microphone)</td>
<td>- Different views providing unique information about length, movement and extensibility of the velum, forward movement of the posterior pharyngeal wall (mid-sagittal view), lateral wall movements (coronal view), type and extent of velopharyngeal closure (axial view, level hard palate)  - High spatial resolution and superior visualization of soft tissues  - No ionizing radiation, repeatable  - Early assessment of submucous cleft palate</td>
<td>- Time-consuming procedure  - Possible consequences of gravity on the velum  - Expensive  - Reduction of image quality because of fixed intraoral metallic prostheses</td>
</tr>
<tr>
<td>Lateral cephalometric radiographic analysis</td>
<td>Lateral cephalograms by x-rays and tracing of the radiographs by using specific software programs</td>
<td>- Standardized information  - Minimal patient compliance  - Allocation of the soft tissues of the nasopharynx to the bony landmarks of the face and the cranium</td>
<td>- Ionization radiation  - Interpretation can be difficult because of multiple shadows  - Limited, static information about the physiology of the velopharynx, only in the midsagittal plane  - Two-dimensional representation of the three-dimensional anatomy</td>
</tr>
<tr>
<td>Computed tomography</td>
<td>Computerized tomography scan (CT scan) on the axial plane</td>
<td>- Determination of velopharyngeal closure level referring to organs  - Quantifying surfacial and deep craniofacial structures  - Promising 3D and 4D imaging</td>
<td>- Ionization radiation  - Limited, static information about the physiology of the velopharynx, only in the axial plane  - Two-dimensional representation of the three-dimensional anatomy</td>
</tr>
</tbody>
</table>
### Appendix B – (continued)

<table>
<thead>
<tr>
<th>Technique</th>
<th>Procedure</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Direct assessment techniques</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Ultrasound                       | Transducer placed against the neck, under the ear and behind the ramus of the mandible in combination with an acoustic coupling gel | - No ionizing radiation  
- Noninvasive, repeatable                                                    | - Limited visualization because of movements of the transducer and signal reflection  
- Intersubject and interobserver variability                                 |
| **Indirect assessment techniques** |                                                                           |                                                                            |                                                                            |
| Acoustic techniques              |                                                                           |                                                                            |                                                                            |
| Horii's Oral Nasal Coupling (HONC index) | - Nasal and laryngeal accelerometer on nose and throat resp. or nasal accelerometer and microphone  
- Determination of constant k  
- Ratio of nasal to vocal accelerometer signal amplitude | - Ratio reduces influence of vocal intensity  
- Inter- and intraspeaker comparisons possible  
- Moderate to strong correlation with perceptual evaluation  
- Simple, noninvasive  
- Good interobserver reliability | - Limited application  
- Not commercially available as preassembled package |
| Nasality Oral Ratio Meter (NORAM) | - Nasal and laryngeal accelerometer on nose and throat resp.  
- Ratio of nasal to vocal accelerometer signal amplitude | - Examination of hypernasality before and after surgery  
- Good correlation with perceptual judgment  
- Noninvasive, convenient, easy to interpret  
- Good test-retest reliability  
- Extensive clinical and research application  
- Quantitative  
- Can be used during feedback | - Low inter- and intraobserver reliability  
- Impossible to distinguish normal resonance from hypernasality  
- Limited application  
- Low correlation with perceptual judgment  
- Use of different cutoff scores to determine sensitivity and specificity of the Nasometer |
| Nasometry                        | - Fixation of a plate on a headset between upper lip and nose  
- Two microphones collect the nasal and oral signal  
- Filtering of the signals through 500Hz filter with 300Hz bandwidth  
- Calculation of nasalance score, range 0-100% | - Good correlation with perceptual judgment  
- Noninvasive, convenient, easy to interpret  
- Good test-retest reliability  
- Extensive clinical and research application  
- Quantitative  
- Can be used during feedback | - Limited application  
- Low correlation with perceptual judgment  
- Use of different cutoff scores to determine sensitivity and specificity of the Nasometer |
### Appendix B – (continued)

<table>
<thead>
<tr>
<th>Technique</th>
<th>Procedure</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acoustic techniques (continued)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spectral characteristics</td>
<td>Digital processing techniques based on specific spectral characteristics of hypernasal speech: pole-zero pairs in the region of F1, reduction of amplitude of F1 and F2, increase of bandwidth of F1 and F2, shifts in formant frequencies, rise in amplitude between F1 and F2</td>
<td>- Techniques can be used as screening instrument</td>
<td>- Features can only be extracted from vowels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Quantitative, noninvasive, cost-effective</td>
<td>- No degree of hypernasality is determined</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Limited need for equipment: microphone, computer and free software</td>
<td>- Difficult to verify the effect of therapy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Influencing variables</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Low correlations with nasalance scores</td>
</tr>
<tr>
<td><strong>Aerodynamic techniques</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal and oral airflow</td>
<td><strong>Mirror-fogging test by Glatzel</strong> - Holding a cold mirror 0.5cm under the nose during phonation of vowels or consonants - The degree of condensation is represented by 4 concentric circles by which 0 corresponds with no condensation and 4 to severe condensation</td>
<td>- Easy to use, simple training - Inexpensive, rapid, noninvasive</td>
<td>- Low reliability and validity - Tilting errors - Influencing variables</td>
</tr>
<tr>
<td></td>
<td><strong>Aerophonoscope</strong> - Three airflow sensors, one for each nostril and one for oral airflow, enable the simultaneous visualization of both, nasal and oral airflow as well as voice levels</td>
<td>- Continuous speech possible - Determination of nasal emission - User-friendly, portable - Can be used during feedback training</td>
<td>- Impairment of speech movements possible because of positioning of the handpiece in front of the mouth</td>
</tr>
<tr>
<td></td>
<td><strong>Pneumotachograph</strong> - Measuring the amount of nasal airflow by a pneumotachograph attached to a nasal continuous positive airway pressure mask</td>
<td>- Safe and easy to perform - Noninvasive - High sensitivity and specificity - Good correlation with perceptual evaluation</td>
<td>- No direct quantification of the degree of hypernasality and movements of the velopharyngeal mechanism - Possible influence on the sensory feedback of speech because of the nasal mask</td>
</tr>
</tbody>
</table>
### Aerodynamic Techniques (continued)

<table>
<thead>
<tr>
<th>Technique</th>
<th>Procedure</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasal and oral airflow</td>
<td><strong>Super Nasal Oral Ratiometry System (SNORS)</strong></td>
<td>- Quantitative, inexpensive and noninvasive</td>
<td>- Influencing variables</td>
</tr>
<tr>
<td>(continued)</td>
<td>- Modified oxygen mask including airflow sensors and microphones</td>
<td>- Information regarding timing of the velopharyngeal closure</td>
<td>- Possible influence on the sensory feedback of speech because of the nasal mask</td>
</tr>
<tr>
<td></td>
<td>- High-speed sensors detect sudden changes in airflow caused by rapid movement of the velum (heated thermistor techniques)</td>
<td>- Can be used during feedback training</td>
<td></td>
</tr>
<tr>
<td>Pressure-flow technique</td>
<td>- Two flexible catheters (within the mouth and nostril) collect intraoral and nasal air pressure, transmitting it to pressure transducers</td>
<td>- Information about timing characteristics associated with speech</td>
<td>- No clarity about the correction factor</td>
</tr>
<tr>
<td></td>
<td>- A pneumotachograph, connected by plastic tubing to the other nostril, measures nasal airflow</td>
<td>- Verifying treatment outcome</td>
<td>- Equipment often not available</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Significant for research</td>
<td>- Technically complex procedures</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Limited significance for daily clinical practice</td>
</tr>
</tbody>
</table>
References


43


Instrumental and perceptual assessment of the velopharyngeal function and resonance


Instrumental and perceptual assessment of the velopharyngeal function and resonance


47


CHAPTER 2

RESEARCH OBJECTIVES OF THE PRESENT STUDY REGARDING THE NASALITY SEVERITY INDEX 2.0

For decades, researchers have been searching for the most ideal assessment technique in order to diagnose resonance disorders and to decide on the most appropriate treatment. On the one hand, accurate information based on direct techniques is needed to diagnose the amount of velopharyngeal dysfunction. On the other hand, resonance disorders should be determined by using perceptual assessment and indirect techniques. As the ideal technique, as described by Horii (1980), has not been found, treatment decisions should not be based on a single source of information on patient performance.

Determining the presence and degree of resonance disorders is essential during diagnosis and/or treatment evaluation as this influences further therapy policies. Although the visualization of the anatomy and physiology of the velopharyngeal mechanism can provide additional information, the relationship between anatomic and physiologic deviations and resonance disorders is not always straightforward (Baken & Orlikoff, 2000; Kummer et al., 2003; Lam et al., 2006; Witt & D'Antonio, 1993). Furthermore, the presence of resonance disorders can influence speech intelligibility (De Bodt et al., 2002) and acceptability (Henningsson et al., 2008; Whitehill et al., 2011) which is often considered the main outcome for surgery and speech therapy (Henningsson et al., 2008; Sell, 2005). As discussed above, resonance disorders can be assessed by using perceptual judgments and indirect assessment techniques. However, reliability and validity issues often occur in perceptual judgments and the perfect indirect assessment technique, which closely reflects the capabilities of the human ear, is not yet available.

A possible solution to sidestep the limitations of single indirect instrumental assessment techniques is the combination of different variables into a multiparametric index that allows to implement complementary indirect measurements. In this light, the initial aim of this doctoral thesis was to explore the application of a Nasality Severity Index (Van Lierde et al., 2007) as a new, multiparametric approach to determine hypernasality in daily clinical practice. To achieve this objective, the following research aims were formulated:

A large standard deviation was found for the original NSI of the control group in the pilot study by Van Lierde et al. (2007). Therefore, the first aim of this doctoral thesis was to collect NSI data in a larger control group with a wider age range to explore the possible influence of personal and environmental variables on the NSI. Additionally, the availability of reference values for Dutch-speaking Flemish children without resonance disorders between 4 and 12 years old was aimed. Based on literature, a(n) (limited) effect of age and gender on the NSI was hypothesized. The results of this study are discussed in chapter 3.
Based on the above-mentioned study, large interindividual differences were found for the original NSI, which prevented the original NSI to be a reliable instrument to assess the hypernasality of an individual person. Moreover, possible influence of personal variables on the parameter ‘maximum duration time of /s/’, such as vital capacity (Mendes Tavares et al., 2012; Tait et al., 1980), language and dental anomalies (Campbell et al., 2008), and the influence of environmental variables on the mirror-fogging test by Glatzel, more specifically air moisture, temperature and tilting errors (Brescovici & Roithmann, 2008), were the main drawbacks. Therefore, an adaptation of the original NSI to the NSI version 2.0 was aimed, based on the inclusion of other predictors for the statistical discrimination of children with perceived hypernasality and control children without resonance disorders. As hypernasality affects speech intelligibility and acceptability more than hyponasality does (Shprintzen et al., 1979), the purpose of the NSI 2.0 was to identify hypernasality, so no specific nasal stimuli were included in the regression analysis. Additionally, the new index was validated based on the data of an independent group of children with perceived hypernasality and control children without resonance disorders. The derivation of the Nasality Severity Index 2.0 is described in chapter 4.

One of the necessary conditions to implement this new index in daily clinical practice and to evaluate interventions properly, is the availability of normative values derived from children and adults without resonance disorders. To formulate these reference values, the possible influence of gender and age on the NSI 2.0 was explored to verify the need for separate reference values according to gender and/or age. Therefore, the third aim of this thesis was to formulate reference values for the NSI 2.0 in normal developing, Dutch-speaking Flemish children and adults without resonance disorders. Some influence of gender and age was hypothesized based on literature, in which women may show lower NSI 2.0 scores compared to men and children may show higher NSI 2.0 scores compared to adults. The results of this study are discussed in chapter 5.

A second condition that has to be verified before implementing a new instrument into daily clinical practice is the reliability of this instrument. Hence, the fourth aim was to explore the short-term and long-term test-retest reliability of the NSI 2.0 in children and adults. The long-term variability of the NSI 2.0 and its parameters was hypothesized to be larger compared to the short-term variability. Additionally, the variability of the NSI 2.0 and its parameters was expected to be larger in children than in adults. This topic is handled in chapter 6.

The final purpose of the NSI 2.0 was to provide an easy-to-interpret severity score of hypernasality to facilitate the evaluation of therapy outcomes, communication to the patient and other clinicians, and decisions for treatment planning. Therefore, the fifth aim was to determine the relationship between the NSI 2.0 scores and perceptual judgments of hypernasality based on ratings of an expert panel. A negative correlation between the NSI 2.0 and perceived severity of hypernasality was hypothesized in which a more negative NSI
2.0 score correlates with the perception of more severe hypernasality. Moreover, due to its multidimensional approach, higher correlations with auditory perceived hypernasality ratings were hypothesized in comparison with a single parameter approach. Additionally, the possible influence of audible nasal airflow and speech intelligibility on the index scores was investigated, in which no influence of these variables was expected. This information is presented and discussed in chapter 7.

Finally, the NSI 2.0 was applied in a clinical study as an indirect measurement to verify the effect of short intensive speech therapy on the degree of hypernasality in five children with a history of cleft (lip and) palate. It was hypothesized that the NSI 2.0 can be applied to indicate the effect of speech therapy on hypernasality. Additionally, the effect of this intensive speech therapy on articulation was verified using narrow phonetic transcription of both words and sentences. The results of this study are described in chapter 8.
References


CHAPTER 3

EFFECTS OF AGE AND GENDER IN NORMAL-SPEAKING CHILDREN ON
THE NASALITY SEVERITY INDEX: AN OBJECTIVE MULTIPARAMETRIC
APPROACH TO HYPERNASALITY

Based on: Bettens, K., Wuyts F.L., De Graef C., Verhegge L., & Van Lierde K.
(2013) Effects of age and gender in normal-speaking children on the Nasality Severity Index:
an objective multiparametric approach to hypernasality. Folia Phoniatica et Logopaedica,
65, 185-192.

Abstract

Objective. Since resonance disorders have a multidimensional nature and occur in several
craniofacial pathologies, the aim of the present study was to determine the influence of age
and gender in normal-speaking children without resonance disorders on an objective
multiparametric index of hypernasality.

Patients and Methods. A total of 74 Dutch-speaking Flemish children (37 boys and 37 girls),
aged 4-12 years, without resonance disorders were included. A Nasometer was used to
obtain nasalance scores (1 vowel, 2 reading passages). An aerodynamic value was calculated
using the maximum duration time and the mirror-fogging test by Glatzel was applied to
visualize nasality as measured by condensation. With the obtained results, a ‘Nasality
Severity Index’ (NSI) was calculated.

Results. A significant age effect was found, in which the NSI increased with increasing age
(p<0.001). No significant difference for the NSI was detected concerning gender (p>0.05).
Unfortunately, considerable standard deviations of the mean NSI were found.

Conclusion. Although a multiparametric index can form a more powerful approach in the
assessment of and treatment planning for individuals with hypernasality, the present study
revealed large interindividual differences in the current NSI. Therefore, adaptation of the
current NSI is recommended, by which the influences of personal and environmental
variables are taken into account.

Key words. Nasality Severity Index; hypernasality
Chapter 3

Introduction

Resonance disorders are a multidisciplinary problem and occur in multiple craniofacial pathologies such as cleft palate, craniofacial disproportions, as well as after adenotomy, after maxillary surgery or in subjects with functional velopharyngeal problems. To determine the degree of resonance disorders, several specialists rely on perceptual measurements as well as on quantitative measurements of aerodynamic and acoustic characteristics. According to several authors (Keuning et al., 2002; Lewis et al., 2003; Nellis et al., 1992; Prado-Oliveira et al., 2015; Shprintzen & Bardach, 1995; Van Lierde et al., 2007; Watterson et al., 1993), these subjective and quantitative measurements have to be interpreted with care and can lead to contradictory results when assessing the resonance of an individual person. This suggests the need for a quantitative measurement for determining nasal resonance disorders by using a multivariate approach. The Nasometer, developed by Fletcher and Bishop (1973) and manufactured by Kay Elemetrics (NJ, Lincoln Park), seems to be an effective instrument for detecting patients with perceived hypernasality in their speech (Dalston et al., 1993; Sweeney & Sell, 2008). However, several authors (Dalston et al., 1993; Keuning et al., 2002) have indicated the risk in determining nasality based on a single assessment protocol and suggested a combination of both instrumental and perceptual judgment. Moreover, Van Lierde et al. (2007) reported a large overlap between some isolated nasalance values when groups of subjects with normal resonance and slight hypernasality were compared. Furthermore, perceptual judgments do not always agree with instrumental assessment (Keuning et al., 2002; Lewis et al., 2003; Watterson et al., 1993). In that case, a combination of several instrumental techniques may suggest a solution. A multiparametric index makes it possible to implement measurements complementary to nasalance values that should be considered in the judgment of hypernasality as recommended by several authors (Dalston et al., 1991; Seaver et al., 1991; van Doorn & Purcell, 1998). Furthermore, multidimensional indexes have already proved their benefit in clinical practice (the Body Mass Index [weight (kg)/length² (m)] is a good illustration of the efficacy of combining variables). In voice research, multivariate techniques have also proved their usefulness, as shown by the Dysphonia Severity Index (DSI) (Wuys et al., 2000) and the Voice Range Profile Index for Children (VRPiC) (Heylen et al., 1998).

Van Lierde et al. (2007) took the first step in creating a multiparametric protocol to assess nasal resonance disorders by constructing a ‘Nasality Severity Index’ (NSI). The index consists of a combination of both acoustic and aerodynamic parameters. The five parameters used in the NSI calculation are the nasalance values of the vowel /a/, as well as an oral and oronasal passage determined by the Nasometer (Fletcher & Bishop, 1973), the mirror-fogging test by Glatzel (during the production of /a/) (Foy, 1910) and the maximum duration time (MDT; consonant /s/). The NSI equation yields $\text{NSI} = -60.69 - (3.24 \times \text{nasalance oral text (\%)} - (13.39 \times \text{Glatzel value /a/}) + (0.244 \times \text{MDT /s/ (s)}) - (0.558 \times \text{nasalance /a/ (\%)}) + (3.38 \times \text{nasalance oronasal text (\%)})$. The average NSI for children without resonance disorders is +14.9 (SD 16), whereas the average NSI in children with a perceived slight hypernasality is
Effects of age and gender in normal-speaking children on the NSI

-24.0 (SD 21.9). The more negative the NSI value, the stronger the presence of hypernasality. Daily clinical use of the NSI has shown it to correspond well to perceptual evaluation of speech and to be a practical and efficient tool for describing the presence of hypernasality as reported by Van Lierde et al. (2007).

One of the necessary conditions to implement this index in daily clinical practice and to evaluate intervention properly is the availability of normative values derived from children without resonance disorders. In literature, a controversy exists about the differences in nasalance values obtained by the Nasometer (Fletcher & Bishop, 1973), for age and gender (Table 3.1). Some authors report low correlations between age and nasalance (Karakoc et al., 2013; Luyten et al., 2012; Seaver et al., 1991; Van de Weijer & Slis, 1991; Van der Heijden et al., 2011; van Doorn & Purcell, 1998; Whitehill, 2001), while others mention significantly lower nasalance values in younger children compared to older children for nasal sentences (Brunnegard & van Doorn, 2009; Prathanee et al., 2003) and oral texts (Prathanee et al., 2003) or lower nasalance scores for certain speech stimuli in children compared to adults (Abou-Elsaad et al., 2012; Hirschberg et al., 2006; Rochet et al., 1998; Van Lierde et al., 2003). Possible explanations for these differences are the stronger evidence of coarticulation in adult speakers as a result of speaking experience (Thompson & Hixon, 1979), changes in lymphoid and gland tissue at the velopharyngeal port (Becker et al., 2009), and the growth of the oropharynx (Taylor et al., 1996) and nasal cavity (Warren et al., 1990).

Regarding gender, opinions about its correlation with nasalance also vary. Females appear to present significantly greater nasalance scores than males, especially on speech samples including nasal phonemes (Rochet et al., 1998; Van Lierde et al., 2003). Possible explanations for these findings are the underlying anatomical and physiological differences between men and women related to the velopharyngeal closure (Kuehn, 1976; Seaver et al., 1991) or more anticipatory coarticulation in women, resulting in more nasal airflow during the production of oral vowels preceding a nasal consonant (Thompson & Hixon, 1979). Furthermore, an existing mismatch in sensitivity of the two microphones of the Nasometer for certain fundamental frequencies has been proposed to explain minimal differences between males and females (Zajac et al., 1996). Because fundamental frequency evolves with age (de Bodt et al., 2008; Maturo et al., 2012), this mismatch in sensitivity can hypothetically also be responsible for the reported influence of age on nasalance scores. However, the reported differences are small and could be due to intrasubject variability (Brunnegard & van Doorn, 2009). Additionally, several authors (Hirschberg et al., 2006; Kavanagh et al., 1994; Litzaw & Dalston, 1992; Mayo et al., 1996; Nichols, 1999; Okalidou et al., 2011; Tachimura et al., 2000; Van de Weijer & Slis, 1991) found no significant differences between men and women. Between boys and girls with a maximum age of 13 years old, also comparable nasalance values have been reported (Brunnegard & van Doorn, 2009; Luyten et al., 2012; Nichols, 1999; Prathanee et al., 2003; Sweeney et al., 2004; Van de Weijer & Slis, 1991; Van der Heijden et al., 2011; van Doorn & Purcell, 1998; Van Lierde et al., 2003).
Table 3.1 Summary of literature about normative nasalance values and influence by gender and age.

<table>
<thead>
<tr>
<th>Study</th>
<th>Language</th>
<th>Gender M/F, n</th>
<th>Age, years</th>
<th>Speech sample</th>
<th>Effect of gender</th>
<th>Effect of age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abou-Elsaad et al. (2012)</td>
<td>Arabic (EGY)</td>
<td>92/76/132</td>
<td>3-9/9-18/18-54</td>
<td>SNAP test</td>
<td>Tendency for increased values in females in syllables and vowels, but not in sentences</td>
<td>Tendency for lower values in children (except in /mi/ syllables): Plosive sentences: 21%±8 and 19%±10 vs. 32%±16 Fricative sentences: 15%±6 and 18%±9 vs. 23%±10</td>
</tr>
<tr>
<td>Van Lierde et al. (2003)</td>
<td>Dutch (BE)</td>
<td>18/15/28</td>
<td>7-13/19-27</td>
<td>Vowels /a/ /i/ /u/ /m/ Oral text Oronasal text Nasal sentences</td>
<td>No sign. differences between boys and girls; Sign. higher scores in females for /u/ (7.1±1.1 vs. 11.8±1.4), oronasal (31.5±0.8 vs. 36.1±0.9) and nasal text (54.2±0.8 vs. 57.4±0.8)</td>
<td>Sign. lower values in children for oronasal (31.9±0.8 vs. 33.8±0.9) and nasal text (51.6±0.8 vs. 55.8±0.8)</td>
</tr>
<tr>
<td>Van de Weijer and Slis (1991)</td>
<td>Dutch (NL)</td>
<td>10/10/10</td>
<td>7-10/20-30</td>
<td>Oral text Oronasal text Nasal sentences</td>
<td>No sign. differences between boys and girls; males and females</td>
<td>No sign. differences between children and adults</td>
</tr>
<tr>
<td>Van der Heijden et al. (2011)</td>
<td>Dutch (NL)</td>
<td>16/19/20</td>
<td>4-5/5-6/6-7</td>
<td>Oral sentences Oronasal sentences</td>
<td>No sign. differences between boys and girls</td>
<td>No sign. effect of age</td>
</tr>
<tr>
<td>van Doorn and Purcell (1998)</td>
<td>English (AUS)</td>
<td>123/122</td>
<td>4-9</td>
<td>Oral text Nasal sentences</td>
<td>No sign. differences between boys and girls</td>
<td>No sign. effect of age</td>
</tr>
<tr>
<td>Rochet et al. (1998)</td>
<td>English (CAN)</td>
<td>149/166</td>
<td>9-85</td>
<td>Oral text Oronasal text Nasal sentences</td>
<td>Sign. higher nasalance values in females for oronasal text (32.9±5.3 vs. 34.5±4.6)</td>
<td>Sign. lower values in children for oral and oronasal text</td>
</tr>
<tr>
<td>Sweeney et al. (2004)</td>
<td>English (IRE)</td>
<td>34/36</td>
<td>5-13</td>
<td>Occlusive sentences Fricative sentences Nasal sentences</td>
<td>No sign. differences between boys and girls</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>
Table 3.1 (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Language</th>
<th>Gender – M/F, n</th>
<th>Age, years</th>
<th>Speech sample</th>
<th>Effect of gender</th>
<th>Effect of age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luyten et al. (2012)</td>
<td>English</td>
<td>35/34</td>
<td>2-13</td>
<td>SNAP test</td>
<td>No sign. differences between boys and girls</td>
<td>No sign. effect of age</td>
</tr>
<tr>
<td>Seaver et al. (1991)</td>
<td>English</td>
<td>56/92</td>
<td>16-63</td>
<td>Oral text Oronasal text Nasal sentences</td>
<td>Sign. higher values in females for nasal sentences (61%±6 vs. 63%±6)</td>
<td>Weak correlation between age and values of oral text (r=0.311)</td>
</tr>
<tr>
<td>Haapanen (1991)</td>
<td>Finnish</td>
<td>5/37</td>
<td>3-54</td>
<td>Oral sentence Oral sentence with plosives Nasal sentence</td>
<td>Not applicable</td>
<td>Decreasing values with rising age for oral sentence with plosives (r=-0.357)</td>
</tr>
<tr>
<td>Rochet et al. (1998)</td>
<td>French</td>
<td>60/93</td>
<td>9-85</td>
<td>Oral text Oronasal text Nasal sentences</td>
<td>Sign. higher nasalance values in females for oronasal text (26.0%±5.2 vs. 28.3%±5.5) and nasal sentences (35.5%±6.7 vs. 38.5%±6.6)</td>
<td>Sign. lower values in children for all stimuli</td>
</tr>
<tr>
<td>Okalidou et al. (2011)</td>
<td>Greek</td>
<td>40/40</td>
<td>18-34</td>
<td>SNAP test</td>
<td>No sign. differences between males and females</td>
<td>No applicable</td>
</tr>
<tr>
<td>Hirschberg et al. (2006)</td>
<td>Hungarian</td>
<td>75</td>
<td>5-25</td>
<td>Vowels Syllables Oral sentence Oronasal sentence Nasal sentence</td>
<td>No sign. differences between males and females</td>
<td>Sign. lower values in children for oral (11.0% vs. 13.4%), oronasal (31.7% vs. 39.5%) and nasal sentence (50.6% vs. 56.0%)</td>
</tr>
<tr>
<td>Tachimura et al. (2000)</td>
<td>Japanese</td>
<td>50/50</td>
<td>19-35</td>
<td>Oral sentences</td>
<td>No sign. differences between males and females</td>
<td>Not applicable</td>
</tr>
<tr>
<td>Nichols (1999)</td>
<td>Spanish</td>
<td>21/26</td>
<td>6-8</td>
<td>Oral sentences Nasal sentences</td>
<td>No relevant differences</td>
<td>No relevant effect of age</td>
</tr>
<tr>
<td></td>
<td>(MEX)</td>
<td>28/22</td>
<td>11-13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24/31</td>
<td>20-40</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.1 (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Language</th>
<th>Gender – M/F, n</th>
<th>Age, years</th>
<th>Speech sample</th>
<th>Effect of gender</th>
<th>Effect of age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brunnegard and van Doorn (2009)</td>
<td>Swedish</td>
<td>21/24</td>
<td>4-6</td>
<td>Oral sentences Oronasal sentences Nasal sentences</td>
<td>No sign. differences between boys and girls</td>
<td>Sign. lower values in younger children for nasal sentences (52.0%±6.6 vs. 56.5%±6.4)</td>
</tr>
<tr>
<td>Prathanee et al. (2003)</td>
<td>Thai</td>
<td>141</td>
<td>6-13</td>
<td>Oral text Oronasal text Nasal sentences</td>
<td>Sign. lower values in girls for the oral text (no separate results reported)</td>
<td>Increasing values with rising age for the oronasal text and nasal sentences</td>
</tr>
</tbody>
</table>

Nasalance values are presented as mean±SD if a significant influence was reported. SNAP=Simplified Nasometric Assessment Procedures, r=Pearson correlation coefficient

Based on this review of the literature, contradictory results have been reported concerning the influence of age and gender on nasalance values. However, a limited effect of these factors on the nasalance parameters of the NSI can be hypothesized in children as a majority of the authors report no influence of age and gender in children under the age of 13 years. Nevertheless, age and gender can possibly have a significant influence on the aerodynamic measurement included in the index, more specifically MDT, and on the mirror-fogging test by Glatzel. The relation between MDT of /s/ and age is well documented (Mendes Tavares et al., 2012; Tait et al., 1980), in which MDT increases with rising age in children due to increasing vital capacity. Additionally, differences in MDT are reported for adult men and women (Gelfer & Pazera, 2006) which can induce an effect of gender on the NSI in adults. However, no gender effect for this parameter is reported in children (Mendes Tavares et al., 2012; Tait et al., 1980). To the best of our knowledge, no studies regarding the influence of age and gender on the mirror-fogging test by Glatzel have been published. Because no (Hirschberg et al., 2006; Luyten et al., 2012; Okalidou et al., 2011; Van Lierde et al., 2003) or only limited (Abou-Elsaad et al., 2012) differences are reported for the nasalance scores of the vowel /a/ considering age and gender (Table 3.1), only limited differences due to these influencing variables are hypothesized for the mirror-fogging test during the production of /a/.

Given these results concerning the influence of age and gender on the included parameters, a (limited) effect of age and gender on the NSI can be hypothesized. If an influence of age or gender is detected, this information will be important in the derivation of normative values for the NSI. Therefore, the purpose of the present study was to determine possible effects of age and gender on the NSI in a group of 102 normal-Dutch-speaking Flemish children between 4 and 12 years old. Because the isolated influence of age and gender on the NSI
cannot be determined in patients with a resonance disorder, only an evaluation of normal-speaking children without resonance disorders is included in the present report.

**Method**

**Participants**

A total of 102 children were invited for this study. They all live in Flanders (the northern part of Belgium) and have Dutch as their mother tongue. These subjects were recruited at random in different (preschool) kindergartens and elementary schools and were selected based on the following criteria: absence of moderate to severe hearing problems, neurological or velopharyngeal problem, absence of a developmental delay including articulation errors, absence of a general disability, orthodontic treatment and oral surgical intervention. Children with a dialect were also excluded from the study. All children’s parents were asked to complete a questionnaire to verify the aforementioned factors. To determine resonance, articulation errors and voice quality two speech-language pathologists (C.D.G and L.V.) judged a 5-min sample of conversational speech of each child on a 2-point rating scale (0=normal resonance/articulation/voice quality, 1=abnormal resonance/articulation/voice quality). Four children were excluded because of a cold resulting in abnormal resonance (n=1), abnormal vocal quality (n=1) or adenotomy (n=2). Ninety-eight children were selected for further investigation; however, six of them were non-cooperative. A total of 92 children completed all assessments successfully, comprising 43 boys and 49 girls. To create equal groups controlled for age and gender, 37 boys and 37 girls were selected randomly and matched for age, ranging from 4 years to 12 years 6 months (mean (SD); boys: 8.3y (2.0); girls: 8.4y (2.2)). Written informed consent was obtained from the parents. This research was approved by the institutional review and ethical board of Ghent University Hospital.

**Materials and procedure**

To assess the parameters that determine NSI values, instrumental assessment techniques were used to determine the nasalance and aerodynamic capacities of the participants in accordance with the protocol of Van Lierde et al. (2007).

**Nasometric values.** To determine nasalance values, the Nasometer (model 6200-3 IBM PC; Fletcher and Bishop (1973), Kay Elemetrics Corporation, NJ, Lincoln Park), was used. The sound separation plate of the headset was fixed between the upper lip and the nose of the child. Two microphones, one superior and one inferior to the plate, captured the oral and nasal signals, which were analyzed to yield a nasalance score. After calibration of the device as described by the manufacturer’s manual (Kay Elemetrics Corporation, 1994), each child was asked to sustain the vowel /a/ and to read or repeat two text passages using a comfortable loudness and pitch. These two passages are similar to those applied in the study by Van Lierde et al. (2007); they were originally developed by Van de Weijer and Slis (1991). The first passage, an ‘oral’ text, consists exclusively of oral speech sounds comparable to the
‘zoo’ passage in English (Fletcher, 1978) and represents most oral speech sounds in Dutch (Van de Weijer & Slis, 1991). The second passage, an ‘oronasal’ text, contains approximately the same percentage of nasal phonemes (11.67%) as in spontaneous Dutch speech (11.63% (Van den Broecke, 1988)) and is similar to the ‘rainbow’ passage (Fairbanks, 1960). Each child was asked to read or repeat each passage once. If the child made a reading error, he/she was asked to read the passage again.

**Aerodynamic measurement.** To determine the aerodynamic values, MDT was used. MDT can be defined as the longest prolongation of a voiceless consonant after maximal inhalation (Van Lierde et al., 2007). The MDT was determined three times by prolonging the /s/ in a sitting position after a model of /s/ production was provided to the child. Only the longest production was used in further investigation. The children were visually stimulated and coached by the investigator during phonation.

To visualize the nasal airflow, the mirror-fogging test by Glatzel (Foy, 1910) was applied. During this test, a cold mirror is held 0.5cm under the nose of the child during phonation of vowels or consonants. The degree of condensation (0-4) is represented by four concentric circles, where 0 corresponds to ‘no condensation’ and 4 to ‘severe condensation’. By producing oral speech sounds, hypernasality can be detected, whereas nasal speech sounds can be applied to determine hyponasality. Following Van Lierde et al. (2007), the children were asked to sustain the oral vowel /a/.

**Statistical analysis.** IBM SPSS Statistics software version 19.0 (IBM Corp., Armonk, NY) was used for statistical analysis of the data. We applied an analysis of covariance (ANCOVA) to investigate the effect of gender on the NSI and its parameters, with age as a covariate. When a gender effect was found without an age effect, an unpaired Student t-test was applied to determine the significance level of difference. When continuous age effects were discovered, a linear regression analysis was performed. A probability level of less than 0.05 was considered to be significant.

**Results**

By comparing both gender groups, no statistically significant age difference was found between boys and girls (unpaired Student t-test: t(72)=0.22, p=0.830). Before proceeding with the ANCOVA, the assumption of homogeneity of regression slopes was ascertained for the NSI (F(1,70)=0.10, p=0.750) and all separate parameters. Because no significant results were found, the assumption was accepted, so that the analysis could be continued.

**Effect of age**

The covariate age was statistically significantly related to the NSI (ANCOVA: F(1, 71)=19.27, p<0.001). Therefore, a linear regression analysis was performed on the NSI with age as the independent variable (F(1, 72)=19.74, p<0.001, R²=0.212). The NSI increased with
Effects of age and gender in normal-speaking children on the NSI

rising age, more specifically by 0.30 points per month or by 3.59 points per year, as illustrated in Figure 3.1. Furthermore, age was also significantly related to the MDT of /s/ (ANCOVA: $F(1, 71)=18.99$, $p<0.001$). A subsequent linear regression analysis with age as the independent variable revealed a statistically significant increase in MDT with age ($F(1,72)=19.18$, $p<0.001$, $R^2=0.210$). The covariate age had no statistically significant influence on the nasalance score of the vowel /a/ (ANCOVA: $F(1,71)=0.40$, $p=0.531$). Additionally, no influence of this covariate was found on the parameters ‘oral text’ (ANCOVA: $F(1,71)=1.64$, $p=0.205$) and ‘oronasal text’ (ANCOVA: $F(1,71)=2.17$, $p=0.146$). For the mirror-fogging test, results for 71 of the 74 children were located in the category ‘no condensation’ and only three children were located in the category ‘light condensation’. Therefore, age effect on this parameter could not be examined statistically.

![Figure 3.1 NSI in function of age and gender, (boys, n=37; girls, n=37). 95% prediction intervals are provided. ANCOVA revealed no statistically significant influence of gender on NSI (p>0.05). However, regression analysis indicated an age effect [NSI = –22.5 + (0.30 × age in months); p<0.001], in which age explains 21% of the variability.

Effect of gender

Table 3.2 summarizes the results for the influence of gender on the NSI and its parameters. Inspection of Table 3.2 shows that no statistically significant gender-related difference was found for NSI when age was considered as a covariate (ANCOVA: $F(1, 71)=0.25$, $p=0.621$). Additionally, no statistically significant gender differences were encountered for the nasalance score of the vowel /a/ (ANCOVA: $F(1, 71)=0.014$, $p=0.905$) and MDT of /s/ (ANCOVA: $F(1,71)=1.26$, $p=0.266$) when the effect of age was controlled for. Concerning the oral and oronasal text, a statistically significant difference between boys and girls was detected after controlling for the effect of age (ANCOVA: $F(1,71)=5.41$, $p=0.023$ and $F(1,71)=5.71$, $p=0.020$ respectively). Because the covariate age had no significant influence on these parameters, as mentioned above, an unpaired Student t-test was performed ($t(72)=-2.35$, $p=0.022$ and $t(72)=-2.33$, $p=0.022$ respectively) in which a significantly higher score was detected for the boys. For the mirror-fogging test by Glatzel of /a/, gender effect could not be examined statistically because of the afore-mentioned skewness of the sample.
Table 3.2 Influence of gender on the Nasality Severity Index (NSI) and its parameters (nasalance score of the vowel /a/, oral text and oronasal text, Maximum Duration Time (MDT) of /s/). The mean, standard deviation of the mean (SD), 95% prediction interval (95%PI) and applied statistical method are provided. As no statistical analysis could be performed on the results of the parameter ‘mirror-fogging test by Glatzel of vowel /a/’ because of extreme skewness, the results of this parameter are not included in the table.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Boys (n=37)</th>
<th>Girls (n=37)</th>
<th>Statistical analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean  SD</td>
<td>95%PI</td>
<td>Mean  SD</td>
</tr>
<tr>
<td>Nasometry (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vowel /a/</td>
<td>7.7  2.08</td>
<td>3.6-11.9</td>
<td>7.9  2.02</td>
</tr>
<tr>
<td>Oral text</td>
<td>15.6 6.07</td>
<td>3.5-27.7</td>
<td>12.6 4.94</td>
</tr>
<tr>
<td>Oronasal text</td>
<td>36.5 6.68</td>
<td>23.2-49.8</td>
<td>33.0 6.33</td>
</tr>
<tr>
<td>Aerodynamic measurement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDT (s)</td>
<td>8.7 1.67</td>
<td>5.4-12.1</td>
<td>10.0 1.74</td>
</tr>
<tr>
<td>NSI</td>
<td>+8.1 13.80</td>
<td>-19.3 - +35.5</td>
<td>+6.6 18.69</td>
</tr>
</tbody>
</table>

*Significant difference (p<0.05)

Discussion

The purpose of the present study was to determine the possible influence of age and gender on the NSI – an instrumental, multiparametric approach to hypernasality – in normal-Dutch-speaking Flemish children between 4 and 12 years of age, to determine normative values for this population. A gradual increase in NSI was encountered with increasing age. This can be explained by the significant increase in MDT of /s/ by age. Because the duration of a voiceless phoneme is limited by the vital capacity (de Bodt et al., 2008), an increase in this capacity, when children become older, can explain this increase in MDT (Mendes Tavares et al., 2012; Tait et al., 1980). Furthermore, it is known that the size, shape and surface of the infraglottic and supraglottic resonating structures and cavities influence vocal resonance (Shprintzen & Bardach, 1995), but before puberty these dimensional characteristics differ very little between boys and girls (Jeans et al., 1981; Vilella et al., 2006). The observation made in our study was that the NSI is not significantly different between boys and girls, which thus corroborates the hypothesis of equal structural anatomy of the nasopharynx and adenoidal development (Jeans et al., 1981; Vilella et al., 2006) and comparable vital capacity (de Bodt et al., 2008) before puberty.

However, large standard deviations of the mean were obtained for the NSI (mean (SD): boys 8.1 (13.80), girls 6.6 (18.69)), which indicate a great spread of NSI values in children without resonance disorders, even in children at the same age (Figure 3.1). Although a large group of children was included in the present study, the interindividual differences remain too large to consider the NSI a reliable instrument for the assessment of the resonance of an
individual person. Therefore, an adaptation of the NSI is necessary to fulfill the aim of this new approach: to provide an instrumental, multiparametric assessment of hypernasality resulting in an unambiguous result to guide treatment planning for individuals presenting with hypernasality.

Based on the results of the current study, age has to be taken into account when MDT is included in the adapted NSI. Additionally, attention must be paid to the influence of personal variables such as vital capacity (Mendes Tavares et al., 2012; Tait et al., 1980), and language and dental anomalies (Campbell et al., 2008) on this parameter. Considering the mirror-fogging test by Glatzel (Foy, 1910), environmental variables such as air moisture, temperature and tilting errors may affect the amount of condensation (Brescovici & Roithmann, 2008). However, in this study 71 of the 74 participants were located in the category ‘no condensation’, which leads to the suspicion of no influence of these variables on the observation. Nevertheless, these influencing variables have to be controlled for in future research and application in daily clinical practice.

Although a multiparametric index can form a new, more powerful approach in the assessment of and treatment planning for individuals with hypernasality, the present study revealed some severe limitations of the current NSI. The large interindividual differences in a large group of children with normal resonance may possibly be due to the influence of personal and environmental variables on the parameters included. Therefore, adaptation of the current NSI with inclusion of new assessment techniques is recommended, in which the influence of personal and environmental variables has to be taken into account.

**Disclosure statement**

None of the authors had any commercial associations, supporting funds or financial disclosures that might pose or create a conflict of interest with information presented in this paper.
References


CHAPTER 4

THE NASALITY SEVERITY INDEX 2.0:
REVISION OF AN OBJECTIVE MULTIPARAMETRIC APPROACH TO
HYPERNASALITY

The Nasality Severity Index 2.0: Revision of an objective multiparametric approach to
hypernasality. The Cleft Palate-Craniofacial Journal, 53(3), e60-e70.

Abstract

Objective. Due to the multidimensional nature of resonance disorders, multivariate
diagnostic assessment is advisable. The Nasality Severity Index (NSI) is based on this point of
view. Because of the influence of personal and environmental variables on the current NSI,
this study aims to refine this index.

Design. Prospective case-control study.

Setting. Tertiary university hospital.

Patients. A total of 42 patients with cleft (lip and) palate and 50 children without resonance
disorders were tested.

Interventions. Resonance was investigated by perceptual as well as instrumental
measurements. A Nasometer was used to score nasalance, and spectral speech
characteristics of a sustained vowel /i/ were determined, among which the voice low tone to
high tone ratio (VLHR). Binary logistic regression analysis was performed to calculate the
optimal index to discriminate patients from control children. Additionally, the validity of the
index was determined based on data from an independent patient and control group.

Results. The NSI 2.0, a weighted linear combination of three variables, can be obtained using
the equation NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%)) – (0.260 x nasalance oral text (%))
– (0.242 x VLHR /i/ 4.47*F0Hz (dB)). The NSI 2.0 has a sensitivity of 92% and a specificity of
100%. Moreover, it has excellent validity (sensitivity 88%, specificity 89%).

Conclusion. The NSI 2.0 discriminates patients from control children with high sensitivity,
specificity and validity. This multiparametric method can offer a more powerful approach in
the assessment of and treatment planning for individuals with hypernasality.

Key words. Assessment; cleft palate; hypernasality; Nasality Severity Index
Chapter 4

Introduction

In patients with cleft lip and palate (CLP), speech improvements are often considered the main outcome of surgical treatment or speech therapy. To determine speech characteristics for the evaluation of surgical and therapeutic outcome in these patients, perceptual judgment is acknowledged as the “gold” standard, although it can be influenced by several variables such as experience (Lewis et al., 2003), type of rating scale (Whitehill et al., 2002), vocal quality (Kataoka et al., 2001), compensatory articulation (Bzoch, 1997) and vowel content (Watterson et al., 2007). Due to this controversy, speech-language pathologists often combine perceptual ratings with instrumental assessment techniques to support their findings. Instrumental techniques can provide information about different aspects of the anatomy and physiology of the included structures or the degree of the resonance disorder. Inclusion of these quantitative measurements allows easy comparison to previously acquired data and contributes to the overall accuracy of the assessment (Vogel et al., 2009).

To determine the degree of hypernasality, acoustic measurement techniques, such as the Nasometer (Fletcher & Bishop, 1973), are available to complement perceptual ratings of resonance. The Nasometer defines the amount of nasalance as the amount of nasal energy in relation to the total amount of nasal-plus-oral energy measured by two microphones on a sound separation plate which is placed under the nose of the patient. The advantages of the Nasometer to evaluate nasalance have been extensively described (Baken & Orlikoff, 2000; Hardin et al., 1992; Karnell, 2011; Sweeney & Sell, 2008; Van Lierde et al., 2001; Watterson & Lewis, 2006). Although the Nasometer is an expensive device, it is being used in different clinical centers and has a broad clinical and research application (Shprintzen & Marrinan, 2009; Stelck et al., 2011; Vijayalakshmi et al., 2007).

Recently, spectral analyses have been described to determine acoustic correlates of hypernasality. Kataoka et al. (1996; 2001) calculated the mean power level of one-third octave intervals with center frequencies from 250 to 8000Hz in children (Kataoka et al., 1996) and from 125 to 6300Hz in adults (Yoshida et al., 2000). One-third octave spectra are applied because they closely represent the frequency resolution of the human ear (Pols et al., 1969). Analyses were performed on a 100ms sample of a sustained vowel /i/ after Fast Fourier Transformation (FFT). The analysis showed an increase of the power level between the first and second formant and decreased amplitudes around the second and third formant regions in samples of patients with perceived hypernasality compared to samples of a small control group without resonance disorders (Kataoka et al., 1996; 2001). Further research in adults with hypernasality due to a range of etiologies by Lee et al. (2003a) was based on the analysis of the vowel /i/ in two non-nasal words. This study revealed that hypernasality is indicated by higher amplitudes at the bands centered at 630, 800 and 1000Hz and by decreased power levels at the band centered at 2500Hz.
The voice low tone to high tone ratio (VLHR) developed by Lee et al. (2003b) is another spectral analysis to detect acoustic correlates of nasal resonance. This technique is based on the introduction of pole-zero pairs due to the coupling of the nasal tract (Lee et al., 2009). As a result of this coupling, an extra ‘nasal’ formant, the pole, appears in the neighborhood of the first formant and the inclusion of the nasal sinuses as a resonance area is assumed to create anti-resonance, the zero (Feng & Castelli, 1996). A power ratio with a specific cutoff frequency between this pole and zero can therefore measure the amount of coupling of the nasal tract (Lee et al., 2009; Tsai et al., 2012). The analysis was first performed in subjects with nasal blockage as a measure of hyponasality using the sustained nasal consonant-vowel syllable /m♯/. After determination of the power spectrum of the sound wave by using FFT, the spectrum was divided into a low-frequency and high-frequency band using a specific cutoff frequency derived from the fundamental frequency (4.47*F0Hz) (Lee et al., 2003b). To quantify the low-frequency power (LFP), the power ranging from 65Hz to the cutoff frequency was summed; the power of the high-frequency band (HFP) was calculated as the summation of the power ranging from the cutoff frequency up to 8000Hz. VLHR is defined as the power ratio of LFP vs. HFP expressed in decibels. The results revealed a significant increase in VLHR after nasal decongestant treatment. This increase can be explained by the increase in nasal resonance after nasal decongestion, resulting in an increase in LFP due to the creation of a nasal formant and a decrease in HFP due to the appearance of anti-resonance, both spectral characteristics of nasal resonance. As a continuation of this study, Lee et al. (2006;2009) determined the most optimal cutoff frequency for six sustained vowels based on correlations with perceptual ratings of hypernasality. Isolated vowels normally do not include acoustic characteristics of nasal resonance. In hypernasal speech, however, pole-zero pairs appear due to the coupling of the nasal tract, which causes an increase in LFP and a decrease in HFP resulting in higher VLHR scores for hypernasal speech. A cutoff frequency between 600Hz to 800Hz was recommended for all six vowels, based on speech stimuli of eight adults. Tsai et al. (2012) applied VLHR to connected speech, including English and Mandarin reading passages, of ten adults without resonance disorders. VLHR as well as nasometry showed significantly higher scores for the nasal passages (loaded with nasal consonants) compared with the oral passages (without nasal consonants). Additionally, significant correlations between VLHR, nasalance and nasality scores were reported.

Several authors consider perceptual assessment of hypernasality the standard against which instrumental acoustic measurements must be validated (Kent, 1996; Keuning et al., 2004; Moll, 1964; Vogel et al., 2009). However, contradictory results have been reported for the extent to which acoustic measurements correspond with perceptual ratings of nasal resonance. For nasometry, correlation coefficients vary between 0.29 to 0.76 based on nasalance scores of oral text passages determined by the Nasometer, with most authors reporting moderate correlations (Brancamp et al., 2010 – r=0.59 and r=0.63; Brunnegard et al., 2012 – r=0.49-0.76; Dalston et al., 1993 – r=0.73; Keuning et al., 2004 – r=0.54; Lewis et al., 2003 – r=0.29-0.57; Sweeney & Sell, 2008 – r=0.74; Watterson et al., 1993 – r=0.49). Additionally, different cutoff scores are proposed to obtain the highest sensitivity (i.e.
correct identification of hypernasality) and specificity (i.e. correct identification of normal nasalance) to discriminate between persons with and without resonance disorders using the Nasometer (Table 4.1).

Table 4.1 Literature overview of reported cutoff scores, sensitivity and specificity to determine hypernasality using a Nasometer.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Rating scale</th>
<th>Nasometric stimulus</th>
<th>Cutoff</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brancamp et al. (2010)</td>
<td>EAI*</td>
<td>Oral text</td>
<td>22%</td>
<td>0.71</td>
<td>0.73</td>
</tr>
<tr>
<td></td>
<td>DME†</td>
<td>Oral text</td>
<td>22%</td>
<td>0.62</td>
<td>0.70</td>
</tr>
<tr>
<td>Dalston et al. (1993)</td>
<td>EAI</td>
<td>Oral text</td>
<td>28%</td>
<td>0.87</td>
<td>0.86</td>
</tr>
<tr>
<td>Hardin et al. (1992)</td>
<td>EAI</td>
<td>Oral text</td>
<td>26%</td>
<td>0.87</td>
<td>0.93</td>
</tr>
<tr>
<td>Sweeney and Sell (2008)</td>
<td>EAI</td>
<td>Oral sentences (total)</td>
<td>35%</td>
<td>0.83</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High pressure sentences</td>
<td>24%</td>
<td>0.83</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low pressure sentences</td>
<td>28%</td>
<td>0.88</td>
<td>0.78</td>
</tr>
<tr>
<td>Watterson et al. (1998)</td>
<td>EAI</td>
<td>High and low pressure sentences</td>
<td>26%</td>
<td>0.84</td>
<td>0.88</td>
</tr>
<tr>
<td>Watterson et al. (1993)</td>
<td>EAI</td>
<td>Oral text</td>
<td>25.25%</td>
<td>0.71</td>
<td>0.55</td>
</tr>
</tbody>
</table>

*EAI = equal appearing interval scale, †DME = direct magnitude estimation scale

Considering the analyses based on spectral characteristics of hypernasality, high correlations have been found between perceptual ratings and the amplitudes of specific one-third octave bands based on one-third octave spectrum analysis (Katoaka et al., 1996 (r=0.82, peak power level of the sixth, seventh, eighth multiples and of the tenth, eleventh and twelfth multiples in the normalized 1/3-octave spectra), 2001 (r=0.84; peak power level of the normalized 1/3-octave bands on 1, 1.6 and 2.5kHz); Yoshida et al., 2000 (r=0.84, peak power level of the seventh and twelfth multiples in the normalized 1/3 octave spectra)). However, the perceptual assessments were based solely on the sustained vowel /i/ which may be insufficient to determine the overall degree of hypernasality (Kummer, 2011; Sell, 2005). Moderate to high correlations are reported between VLHR and perceptual ratings (Lee et al., 2009: r_{VLHR}=0.54-0.62; Tsai et al., 2012: r_{VLHR}=0.79-0.81). Nevertheless, no significant differences for VLHR between various grades of hypernasality are reported by Vogel et al. (2009). According to Tsai et al. (2012), these results may be attributed to an incorrect application of the analysis. More specifically, Tsai et al. (2012) emphasize the necessity to use a frequency bandwidth of less than 20Hz and the summation of the magnitudes of each frequency before logarithmic transformation to decibels instead of after in the calculation of VLHR.

Spectral analyses are objective, noninvasive and cost-effective (Baken & Orlikoff, 2000; Vijayalakshmi et al., 2007). However, determination of spectral characteristics can be influenced by the presence of noise, the loudness of the speech sample, interspeaker variation and by the quality of the equipment (Katoaka et al., 1996; Vijayalakshmi et al., 2007; Yoshida et al., 2000). Moreover, the analyses are mostly based on vowels only, which may not reflect hypernasality in connected speech.
To sidestep the limitations of these currently advocated single assessment methods, a combination of different assessment techniques may offer a solution. Multiparameter indices have already proved useful in clinical practice (the Body Mass Index [weight (kg)/length² (m)] is a good illustration of the usefulness of combining variables). In voice research, multivariate techniques have also proved their advantage as shown by the Dysphonia Severity Index (DSI) (Wuysts et al., 2000) and the Voice Range Profile Index for Children (VRPlc) (Heylen et al., 1998). Van Lierde et al. (2007) proposed a first step in creating a multiparametric protocol to assess nasal resonance disorders by constructing a ‘Nasality Severity Index’ (NSI). The NSI equation aimed at a statistically optimal discrimination between children with normal resonance versus those with perceived hypernasality due to a history of cleft palate, by means of a weighted combination of both aerodynamic and acoustic measurements. The five parameters applied in the NSI equation are three nasalance scores determined with the Nasometer (Fletcher & Bishop, 1973), more specifically the vowel /a/, an oral and oronasal text passage (Fletcher & Bishop (1973); Van de Weijer & Slis (1991), see appendix), the mirror-fogging test by Glatzel (Foy, 1910) during the production of /a/ and the maximum duration time (MDT) of /s/. The equation is NSI = \(-60.69 - (3.24 \times \text{nasalance oral text (\%)} - (13.39 \times \text{Glatzel value /a/}) + (0.244 \times \text{MDT /s/ (s)}) - (0.558 \times \text{nasalance /a/ (\%)}} + (3.38 \times \text{nasalance oronasal text (\%)}}\). The more negative the NSI value, the stronger the degree of hypernasality. The NSI has a sensitivity of 88% and a specificity of 95% which indicates a clear discrimination between children with normal and deviating nasality (Van Lierde et al., 2007). Although this NSI formed a first step in applying a multivariate approach to hypernasality, some limitations emerged during clinical application of the index. A first cause of concern is the use of the mirror-fogging test by Glatzel (Foy, 1910) to visualize nasality as measured by condensation. Given that different mirror types are in use and because of the impact of several influencing variables, such as environmental temperature, air moisture and the method applied (Brescovici & Roithmann, 2008), the reliability and reproducibility of this device can be doubted (Pochat et al., 2012). Second, MDT of /s/ also forms a cause for concern because it is often difficult for young children to sustain this phoneme properly. Furthermore, the position of the central incisors and the vital capacity of a person can influence the sustained production of /s/ (Campbell et al., 2008). Additionally, this test is based on aerodynamic instead of acoustic measurements which are not directly related to pure resonance effects. Due to this parameter, the NSI increased with rising age as discovered in a group of 74 children between 4 and 12 years old without resonance disorders (Bettens et al., 2013). Consequently, this extra influencing factor has to be taken into account when interpreting the results of the NSI. Finally, the aim of the NSI is to provide a protocol to determine the degree of hypernasality in individual patients. However, high standard deviations of the mean NSI were found in this large control group of 74 children. Considering these remarks, an adaptation of the NSI is recommended by adjusting the choice of instrumental techniques used in the regression analysis. Therefore, the aims of the current study were 1) to create a clinically feasible revised NSI (NSI 2.0) to multiparametrically discriminate persons with perceived hypernasality from individuals.
without resonance disorders and 2) to verify the validity of the new equation by determining NSI 2.0 results in a new, independent group of patients speaking English as a second language.

**Methods**

This research was approved by the institutional review and ethical board of the Ghent University Hospital (EC/2012/049).

**Part 1 – Revision of the NSI**

**Participants**

A total of 80 children with cleft palate attending the Ghent University Hospital Craniofacial Center were invited to participate in this study between March 2012 and May 2014. The patients were selected based on the following criteria: having a repaired cleft palate with or without cleft lip (CP±L), being a native speaker of Dutch, being aged between 4 and 13 years old and having no cognitive or neuro-motor disorders. A total of 42 children, 19 girls and 23 boys (mean age 7.8 years, SD 2.59), fulfilled these criteria and agreed to attend the assessments. Thirteen children had a UCLP, 10 had a BCLP and 19 had a CP. Written informed consent was obtained from the parents.

For the control group, 50 children between 4 and 13 years old were selected randomly from several (preschool) kindergartens and elementary schools. To verify inclusion and exclusion criteria, all children’s parents were asked to complete a questionnaire. The control group children were selected based on the following criteria: being a native speaker of Dutch, living in Flanders (the northern part of Belgium), not suffering from a cold or hoarseness at the moment of testing, absence of moderate to severe hearing problems, neurological or velopharyngeal problems, absence of speech disorders, allergy, developmental delay, general disability, orthodontic treatment and oral surgical interventions. Overall, 25 girls and 25 boys (mean age 8.5 years, SD 2.13) were selected.

**Perceptual assessment**

A 5-minute conversational speech sample was digitally video-recorded in a quiet room using a Sony HDR-SR1 camera with integrated high-quality microphone. To determine the degree of hypernasality, hyponasality, audible nasal emission and nasal turbulence, two speech-language pathologists (K.B. and A.L.), both experienced in the evaluation of resonance disorders, judged the speech samples using the definitions and rating system of the Cleft Audit Protocol for Speech – Augmented (CAPS-A) (John et al., 2006). The degree of hypernasality was judged on a 5-point scale (0 = absent, 1 = borderline, 2 = mild, 3 = moderate, 4 = severe), hyponasality was quantified using a 3-point scale (0 = absent, 1 = mild, 2 = marked), audible nasal emission and nasal turbulence were judged on a 3-point scale (0 = absent, 1 = occasionally heard, 2 = frequently heard). Initially, all samples were
independently judged by each investigator. Subsequently, both scores were compared and discussed until a consensus could be reached.

**Instrumental assessment**

Instrumental assessment techniques to determine acoustic characteristics of speech were explored to revise the parameters of the NSI. The techniques were selected based on their ease of use, degree of invasiveness, reproducibility and availability. Consequently, spectral analysis, more specifically 1/3 octave spectrum analysis and VLHR, and nasometry were chosen.

Considering the 1/3 octave spectrum analysis, different frequency bands are reported in the literature that discriminate between normal and hypernasal speech (Kataoka et al., 1996; 2001; Lee et al., 2003a), in which no consensus is yet available about which frequency bands discriminate significantly. Inclusion of the relative amplitude of all 1/3 frequency bands between 200 and 8000Hz, however, is not appropriate in the logistic regression analysis to derive the NSI 2.0. Nevertheless, inclusion of a selection of frequency bands can induce the risk that the acoustic correlates of hypernasality do not occur in those selected frequency bands in each patient. Consequently, only VLHR was included as a prediction variable in the regression analysis. VLHR is based on the same principle as 1/3 octave spectrum analysis, namely a shift of energy due to the coupling of the nasal tract in hypernasal speech, but integrates the full spectrum between 65 and 8000Hz which limits the risks of the selection of a restricted part of the spectrum.

To determine VLHR, each child was asked to sustain the vowel /i/ for at least 2 seconds in front of a unidirectional condenser microphone (Samson, C01U) placed at 10cm from the subject’s mouth. The vowel /i/ was opted for on account of its high velar position which allows the detection of nasal sound transmission even at small degrees of nasal coupling (Fant, 1970). Recordings were made using PRAAT-software, version 5.3.78 (Boersma and Weenink, 2014) at a sampling frequency of 44100Hz. The samples were analyzed based on a 0.5s fragment, extracted using a Hamming window. Two cutoff frequencies were used to determine VLHR: 4.47*F0Hz (Lee et al., 2003b) and 600Hz (Lee et al., 2009). Based on these cutoff frequencies, the spectrum was divided into a low frequency (65Hz to cutoff) and a high frequency (cutoff to 8000Hz) spectral region. After summation of the power of each frequency region, VLHR was determined as the power ratio of the low frequency to the high frequency spectrum in accordance with Lee et al. (2003b) and Lee et al. (2006;2009). The cutoff based on the fundamental frequency was included because formant frequencies decrease with age due to the growth of the vocal tract and resonance areas (Flipsen & Lee, 2012; Vorperian & Kent, 2007). As determination of the cutoff frequency of 600Hz was based on adults (Lee et al., 2009), this cutoff may be insufficient to detect energy changes in the spectrum of children. The fundamental frequency, however, generally depends on body structure (Vorperian & Kent, 2007) and is therefore indirectly related to the formant frequencies.
A Nasometer II model 6450 (KayPentax, NJ, Lincoln Park) was used to quantify nasalance values in three vowels and two text passages. After calibration of the device as described in the manufacturer’s manual (Kay Elemetrics Corporation, 2010) in a quiet room, each child was asked to sustain the vowels /a/, /i/ and /u/ and to read or repeat two passages using a comfortable loudness and pitch. If the child made a reading error, he/she was asked to read the passage again. The first passage, a so called ‘oral’ text, exclusively consisted of oral speech sounds and is normally used to detect hypernasality. The second passage, the ‘oronasal’ text, contained approximately the same percentage of nasal phonemes (11.67%) as in spontaneous Dutch speech (11.63%, Van den Broecke, 1988)). The passages were originally developed by Van de Weijer and Slis (1991) and are available in the appendix.

Statistical analysis

SPSS statistics version 22.0 (IBM Corp., Armonk, NY) was used for the statistical analysis of the data. To compare the results of the individual instrumental assessments from the patient and control group, independent Student t-tests were used with probability levels below 0.05 considered to be statistically significant. Binary logistic regression analysis was applied to select the most optimal combination of prediction variables to separate the control group from the patients perceptually judged with hypernasality.

Part 2 – Validation of the NSI 2.0

Participants

An independent group of 24 Ugandan patients with CP±L (9 boys, 15 girls; mean age 8.1 years, SD 2.94) and 28 children without resonance disorders and born without cleft (13 boys, 15 girls; mean age 8.4 years, SD 1.80) was selected to evaluate the validity of the equation. Children with repaired cleft palate were recruited at the CoRSU hospital (Comprehensive Rehabilitation Services in Uganda) during the VLIR-UOS project of our research unit (ZEIN2009EL28). This project aims to create a reference center for congenital facial clefts and benign jaw tumors in Uganda. All children spoke English as a second language. English is one of the official languages of the country, resulting in a deep embedding of this language in education, media, religion, economics and politics (Mpuga, 2003). The participants were native speakers of languages such as Bantu, Swahili and Nilotic (Mpuga, 2003). Therefore, resonance assessments were performed in their common second language, i.e. English. The same inclusion and exclusion criteria as mentioned in part 1 of this study were applied. Twelve children presented with a UCLP, six had a BCLP and six had a CP. Written informed consent was obtained from the parents. Children without resonance disorders, also speaking English as a second language, were recruited from an orphanage in Uganda. Written informed consent and completed questionnaires were collected from the foster mothers.
Perceptual assessment

A speech sample for perceptual assessment was collected by asking the participants to repeat 12 oral sentences and 3 nasal sentences derived from the MacKay-Kummer Simplified Nasometric Assessment Procedures (SNAP) test (Kummer, 2005; MacKay & Kummer, 1994). Recordings and judgments were made by the same investigators using the same apparatus and rating scales as described in part 1 of this study.

Instrumental assessment

The same test protocol as in part 1 was applied, including a recording of the vowel /i/ to derive spectral characteristics. Furthermore, nasometric values were collected using the same Nasometer II model 6450 (KayPentax, NJ, Lincoln Park) starting from the same type of speech materials, more specifically, the vowels /a/, /i/ and /u/, and an English oral (zoo passage, Fletcher (1978)) and oronasal text (rainbow passage, Fairbanks (1960)).

Statistical analysis

Only those parameters which were determined by the regression analysis in part 1 of this study were used to validate the new index. Additionally, NSI scores were determined for all Ugandan participants by using the NSI 2.0 formula derived in part 1 of this study. Subsequently, the sensitivity and specificity of the NSI 2.0 for this new, independent group were determined.

Results

Part 1 – Revision of the NSI

Perceptual assessment

Table 4.2 outlines the results of the perceptual assessment. Twelve percent (5/42) of the patients were perceptually judged with normal resonance and 5% (2/42) were considered as having mixed nasality. This subgroup (17%, 7/42) was excluded from the data set used to derive the NSI 2.0. Fourteen of the remaining patients (40%, 14/35) were judged with hypernasality and 21 patients (60%, 21/35) were judged with hypernasality in combination with audible nasal emission and nasal turbulence. The purpose of the current study was to develop a multiparametric index to identify patients with hypernasality. Although the influence of audible nasal emission and nasal turbulence on nasalance scores is yet unclear (Sweeney & Sell, 2008), the inclusion of only patients with perceived hypernasality without audible nasal emission and nasal turbulence to derive the equation would limit the clinical application of the index. Therefore, patients with perceived hypernasality whether or not in combination with audible nasal emission and turbulence were included in the data set to derive the NSI 2.0. This resulted in a patient group of 15 girls and 20 boys between 4 years 2 months and 12 years 6 months (mean age 7.6 years, SD 2.25). All children in the control group were judged as having normal resonance.
Table 4.2 Results of the perceptual assessments in the Dutch-speaking Flemish patients with CP±L and the Dutch-speaking Flemish control group.

<table>
<thead>
<tr>
<th></th>
<th>Patients</th>
<th>Absent</th>
<th>Borderline</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypernasality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Hyponasality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Audible nasal emission and turbulence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>50</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Instrumental assessment**

Table 4.3 summarizes the results of the instrumental assessments, i.e. spectral characteristics and nasalance scores. Mean ages did not differ significantly between patients with hypernasality and control children (patient group mean (SD): 7.6y (2.25), control group: 8.5y (2.13), independent Student t-test: t(82) = 1.831, p=0.071). Statistically significant differences between both groups were observed for VLHR of /i/ with a cutoff frequency of 4.47*F0Hz and a cutoff frequency of 600Hz (independent Student t-tests: p<0.05). For all nasalance scores, statistically significant differences were present between the patient and control group (independent Student t-tests: p<0.01).

Table 4.3 Results of the instrumental assessments in the Dutch-speaking Flemish patients with perceived hypernasality and the Dutch-speaking Flemish control group. Means, standard deviations (SD) and ranges are provided.

<table>
<thead>
<tr>
<th></th>
<th>Patient group (n=35)</th>
<th></th>
<th>Control group (n=50)</th>
<th></th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spectral characteristics (dB)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VLHR /i/ (600Hz)</td>
<td>22.88</td>
<td>7.01</td>
<td>3.41-37.21</td>
<td>19.70</td>
<td>4.01</td>
</tr>
<tr>
<td>VLHR /i/ (4.47*F0Hz)</td>
<td>27.82</td>
<td>5.79</td>
<td>14.98-41.05</td>
<td>21.55</td>
<td>4.34</td>
</tr>
<tr>
<td><strong>Nasometry (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a/</td>
<td>24.3</td>
<td>9.11</td>
<td>5-42</td>
<td>9.6</td>
<td>5.94</td>
</tr>
<tr>
<td>/i/</td>
<td>60.5</td>
<td>17.69</td>
<td>14-86</td>
<td>23.7</td>
<td>9.90</td>
</tr>
<tr>
<td>/u/</td>
<td>45.1</td>
<td>17.35</td>
<td>8-77</td>
<td>12.1</td>
<td>5.80</td>
</tr>
<tr>
<td>Oral text</td>
<td>36.9</td>
<td>15.45</td>
<td>9-62</td>
<td>11.2</td>
<td>3.49</td>
</tr>
<tr>
<td>Oronasal text</td>
<td>48.3</td>
<td>11.54</td>
<td>23-71</td>
<td>31.2</td>
<td>5.09</td>
</tr>
<tr>
<td><strong>Nasality Severity Index 2.0</strong></td>
<td>-6.8</td>
<td>5.14</td>
<td>-15.8 - +5.0</td>
<td>+4.1</td>
<td>1.59</td>
</tr>
</tbody>
</table>

*Significant difference, p<0.05; VLHR=voice low tone to high tone ratio, cutoff frequency provided between parentheses.

The predictors for the logistic regression analysis were VLHR of /i/ with a cutoff frequency of 4.47*F0Hz (dB), VLHR with a cutoff frequency of 600Hz (dB), nasalance values of /a/, /i/, /u/, an oral and oronasal text (%), age (years) and gender. A binary logistic regression revealed
The NSI 2.0: Revision of an objective multiparametric approach to hypernasality

The combination of three parameters as an index for hypernasality: the nasalance score of /u/ (%), the nasalance score of an oral text (%) and VLHR of /i/ with a cutoff frequency of 4.47*F0Hz (dB). This resulted in the equation NSI 2.0 = -12.76558 + (0.0824 x nasalance /u/ (%) + (0.260 x nasalance oral text (%)) + (0.242 x VLHR /i/ 4.47*F0Hz (dB)) with a mean NSI score of -3.7 (SD 1.59) for the control group and +7.3 (SD 5.14) for the patient group. Thirty-three out of 35 patients were correctly identified with hypernasality, resulting in a sensitivity of 94% for the NSI 2.0. All children in the control group (50/50) were correctly identified as having no hypernasality, resulting in a specificity of 100%. Analogously to the Dysphonia Severity Index (Wuyts et al., 2000), the presence of hypernasality was preferred to be related to a negative NSI score. Therefore, the signs of the constant and covariates in the equation were switched. To determine the optimal cutoff score, receiver operating characteristic (ROC) curves were conducted. A cutoff score of -0.4327 resulted in a sensitivity of 94% and a specificity of 100% in accordance with the results of the logistic regression analysis. To facilitate the interpretation of the NSI scores, a cutoff score of zero was preferred. Therefore, the constant of the formula was increased by 0.4327 to shift all the scores and to receive a clear cutoff score of zero. These conversions resulted in the final equation of the NSI 2.0, namely NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%)) – (0.260 x nasalance oral text (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)). Based on the beta-coefficients of the predictors (odds ratio [95% CI]: nasalance /u/: 1.086 [0.970-1.215]; nasalance oral text: 1.296 [1.032-1.629]; VLHR /i/: 1.273 [1.014-1.599]), the covariates ‘nasalance of an oral text’ and ‘VLHR of /i/ 4.47*F0Hz’ carry a little more weight in the equation compared to the covariate ‘nasalance of /u/’. More specifically, as the covariates ‘nasalance of an oral text’ or ‘VLHR of /i/ 4.47*F0Hz’ increase by one unit, the change in the odds of being classified as having hypernasality are 1.30 and 1.27 respectively. On the other hand, as the covariate ‘nasalance of /u/’ increases by one unit, the change in the odds of being classified as having hypernasality is 1.09. As a model without ‘nasalance of /u/’ has a sensitivity of 6% less than the current one, this variable still seems to have a discriminating value. The mean NSI 2.0 value of patients with perceived hypernasality is -6.8 (SD 5.14), whereas the mean NSI 2.0 value of the control children with normal nasality is +4.1 (SD 1.59). Zero is the cutoff between normal nasality (NSI 2.0>0) and hypernasality (NSI 2.0<0).

Additionally, 100% (5/5) of the patients who were perceptually judged without hypernasality (and therefore were excluded from the data set to derive the NSI 2.0) were correctly classified as having normal resonance (cutoff>0), resulting in an overall specificity of 100% (55/55) for the total group of children judged without resonance disorders. Of the two patients judged with mixed nasality, one was classified as having normal resonance (NSI 2.0=+3.6, perceptual judgment: mild hypernasality, mild hyponasality) and one as having abnormal resonance (NSI 2.0=-8.4, perceptual judgment: severe hypernasality, mild hyponasality). Addition of those two patients to determine the overall sensitivity results in an overall sensitivity for the NSI 2.0 of 92% (34/37).
Part 2 – Validation of the NSI 2.0

Perceptual assessment

Table 4.4 shows the results of the perceptual assessment of the Ugandan children speaking English as a second language. Twenty-nine percent (7/24) of the patients were judged without hypernasality and were therefore excluded from the analysis, which resulted in a patient group of 8 girls and 9 boys (mean age 9.6 years, SD 3.10). All children of the control group were judged as having normal resonance.

Table 4.4 Results of the perceptual assessments in the Ugandan patients with CP±L and the Ugandan control group, speaking English as a second language.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Absent</th>
<th>Borderline</th>
<th>Mild</th>
<th>Moderate</th>
<th>Severe</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypernasality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients</td>
<td>24</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Controls</td>
<td>28</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Hyponasality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients</td>
<td>24</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>28</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Audible nasal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>emission and</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>turbulence</td>
<td>Patients</td>
<td>17</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Controls</td>
<td>28</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Instrumental assessment

Table 4.5 lists the results of the instrumental assessments, i.e. the results for the distinct parameters and NSI 2.0. Significantly different NSI scores were found (independent Student t-test: \(t(19.7)=8.315, p<0.001\)) between the patient and control group. A sensitivity of 88% (15/17) and a specificity of 89% (25/28) was observed based on the NSI 2.0 for this complete independent validation set using the same cutoff score as for the Dutch-speaking group in part 1 (cutoff=0). Additionally, 86% (6/7) of the patients who were perceptually judged without hypernasality were classified as having normal resonance (cutoff>0), resulting in an overall specificity of 89% (31/35). The patient classified as having abnormal resonance was perceptually judged with occasionally heard nasal turbulence.
**Table 4.5** Results of the instrumental assessments in the Ugandan patients with perceived hypernasality and the Ugandan control group, speaking English as a second language. Means, standard deviations (SD) and ranges are provided.

<table>
<thead>
<tr>
<th></th>
<th>Patients (n=17)</th>
<th>Control group (n=28)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spectral characteristics (dB)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VLHR /i/ (4.47*F0Hz)</td>
<td>27.45, 6.99, 12.87-38.38</td>
<td>20.94, 6.37, 9.34-35.92</td>
<td>0.001*</td>
</tr>
<tr>
<td>VLHR /i/ (600Hz)</td>
<td>20.8, 5.53, 10.92-28.07</td>
<td>20.2, 6.11, 8.26-34.04</td>
<td>0.742</td>
</tr>
<tr>
<td><strong>Nasometry (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a/</td>
<td>25.6, 10.68, 8-47</td>
<td>10.2, 6.84, 4-36</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>/i/</td>
<td>61.6, 19.92, 18-82</td>
<td>22.4, 7.83, 11-39</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>/u/</td>
<td>47.9, 16.40, 18-71</td>
<td>13.2, 6.88, 5-31</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Oral text</td>
<td>46.9, 17.10, 15-75</td>
<td>13.2, 5.67, 6-29</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Oronasal text</td>
<td>52.2, 13.26, 26-78</td>
<td>32.3, 7.00, 23-47</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Nasality Severity Index 2.0</td>
<td>-9.6, 6.13, -20.4 - +1.5</td>
<td>+3.6, 2.50, -4.6 - +7.2</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

*Significant difference, p<0.05; VLHR=voice low tone to high tone ratio, cutoff frequency provided between parentheses.

**Discussion**

Current diagnosis of resonance disorders is based on perceptual as well as quantitative data. However, clinicians are often confronted with contradictory results while examining an individual patient’s resonance (Shprintzen & Bardach, 1995). Therefore, weighted combinations of quantitative assessment results may offer a solution to sidestep the limitations of available, single assessment methods. Following this line of reasoning, Van Lierde et al. (2007) constructed a ‘Nasality Severity Index’, based on a combination of outcomes obtained from different instrumental procedures. However, further clinical application of the index revealed several limitations. Therefore, the current study aimed at a revision of the NSI without the use of the mirror-fogging test by Glatzel (Foy, 1910) or the use of aerodynamic techniques. Furthermore, the validity of this new equation was determined.

The adapted equation yields NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%)) – (0.260 x nasalance oral text (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)). With an overall sensitivity of 92% and an overall specificity of 100%, the NSI 2.0 distinctively identifies Dutch-speaking Flemish children with and without hypernasality. Moreover, the NSI 2.0 recognizes patients with hypernasality whether or not the patient exhibits audible nasal emission or nasal turbulence.

As mentioned above, the Nasometer is being used in different clinical centers and has a broad clinical and research application (Shprintzen & Marrinan, 2009; Stelck et al., 2011; Vijayalakshmi et al., 2007). However, different values of sensitivity and specificity and a variance of cutoff scores are reported in the literature. Based on the results from the oral text recordings in this study, for the nasalance score in itself, an overall sensitivity of 81%
(30/37) and an overall specificity of 93% (51/55) was found using a logistic regression analysis. With an overall sensitivity and specificity of 92% and 100% resp., the NSI 2.0 clearly identifies patients and controls more accurately in comparison with a single parameter approach. Moreover, VLHR implies the analysis of a broader section of the spectrum compared to the bandwidth of 300Hz used by the Nasometer (Fletcher & Bishop, 1973), and the influence of loudness variations is limited by using a relative rather than an absolute index (Lee et al., 2003b). The use of nasalance scores for an oral text also eliminates the restrictions of spectral analysis applied on single vowels. Additionally, an overall sensitivity of 88% and a specificity of 89% were found based on the results of an independent group of Ugandan children speaking English as a second language with and without hypernasality. These results validate the high discriminatory value of the NSI 2.0, both in Dutch and in English, using the same cutoff score to identify children with and without hypernasality.

Eight percent (3/37) of the Dutch-speaking Flemish patients with perceived hypernasality were misjudged as having normal resonance. Two of these patients were perceptually evaluated as having only borderline hypernasality, which may suggest that the perceptual judgments were too strict. Considering that the NSI 2.0 identifies the majority of the patients with borderline hypernasality in this study (78%, 7/9), the index forms an additional, objective appliance in the judgment of resonance disorders, especially for inexperienced speech-language pathologists. The third patient who was misclassified as having normal resonance was perceptually judged with mixed nasality. Although only oral stimuli are included in the NSI 2.0, the combination of mild hyper- and hyponasality could have resulted in a wrong classification because the hypernasality was probably covered by the hyponasality. In contrast, the other patient who was perceptually judged with mixed nasality was correctly classified by the NSI as having hypernasality. However, this patient was judged with severe hypernasality and only mild hyponasality so that the hypernasality was possibly more decisive.

Considering VLHR of a sustained vowel /i/, only Vogel et al. (2009) reported on the comparison of a control and patient group, in which no significant differences were found for VLHR determined with a cutoff frequency of 600Hz. This is in contrast with the findings of the current study in which a significant difference was found for this parameter. However, the VLHR with a variable cutoff frequency of 4.47*F0Hz was selected by the regression analysis as a parameter of the NSI 2.0 equation. A possible explanation is that formant frequencies decrease with age because of the growth of the vocal tract and resonance areas (Flipsen & Lee, 2012; Vorperian & Kent, 2007). Because the extra ‘nasal’ formant in hypernasal speech is situated around the first formant, this energy peak may be included in the high frequency region instead of the low frequency region when a static cutoff score of 600Hz is used. Therefore, a variable cutoff based on the fundamental frequency (which generally depends on body structure and is therefore indirectly related to the formant frequencies (Vorperian and Kent, 2007)) may be more adequate to determine the VLHR in children.
The current mean nasalance values of the sustained vowel /u/ (12.1%) and an oral text (11.2%) of the Dutch-speaking Flemish control group are similar to those reported in literature. Luyten et al. (2012) and Van Lierde et al. (2003) reported a mean nasalance score for a sustained vowel /u/ in children of respectively 17.4% and 9.6%. Moreover, Karakoc et al. (2013), Luyten et al. (2012), Park et al. (2014), Van de Weijer and Slis (1991), van Doorn and Purcell (1998) and Van Lierde et al. (2003) found nasalance scores for an oral text of respectively 15.14%, 14.4%, 11.44%, 11.75%, 13.1% and 11.3%. Although the NSI 2.0 stands a first great test of applicability across languages (Dutch and English), further research is needed to confirm the uniformity of the index in different languages and dialects considering the known influence of language and dialects on nasalance (Brunnegard & van Doorn, 2009; Okalidou et al., 2011; Rochet et al., 1998; Van Lierde et al., 2003).

Furthermore, the possible influence of gender and age has to be explored to determine the need for separate normative NSI 2.0 values for gender and different age groups. As the parameter ‘age’ was not selected by the regression analysis, little influence of age can be expected on the NSI 2.0 in children. This is in contrast with the original NSI in which age was positively correlated to the NSI scores (Bettens et al., 2013). The Pearson correlation coefficient between age and the NSI 2.0 in the control group of the current study is low and non-significant (Pearson correlation coefficient: r=0.21, p=0.138). However, further research is still required to examine the effect of age across a larger age span.

A limitation of this study is that only two speech-language pathologists were responsible for the perceptual assessments. Although other studies also reported a limited amount of raters (54% of the studies reviewed by Lohmander and Olsson (2004) include only one or two listeners), a panel of experienced speech-language pathologists may be preferred to more reliably determine the degree of hypernasality, audible nasal emission and nasal turbulence (Kuehn & Moller, 2000). In further research, the correlation between degree of perceived hypernasality and NSI 2.0 has to be investigated by including perceptual ratings of a larger number of experienced observers. Following this, the aim of grading the degree of hypernasality by the NSI 2.0 can be explored.

In conclusion, the NSI 2.0 is a quantitative identifier of hypernasality based on a multiparametric approach using two different acoustic measurement techniques. This results in an index with high sensitivity and specificity that is noninvasive, easily repeatable, and convenient to establish and interpret. Furthermore, limitations of the original NSI procedure (Van Lierde et al., 2007), such as influence of environmental and personal variables, were restrained. Finally, future research should investigate the value of the NSI 2.0 as a tool to quantitatively evaluate the impact of therapeutic interventions on resonance disorders. In conclusion, the multiparametric NSI 2.0 forms a new, more powerful approach in the assessment of and treatment planning for individuals presenting with hypernasality.
Acknowledgments

The authors gratefully acknowledge all patients and children for their contribution, especially the orphan home Kids of Africa in Entebbe and the kindergartens and elementary schools in Belgium. Furthermore, the authors want to thank Guo-She Lee, Adam Vogel and Alice Lee for the communication and their scripts on VLHR and 1/3 octave spectrum analysis.
References


The NSI 2.0: Revision of an objective multiparametric approach to hypernasality


Chapter 4


**Appendix**

**Oronasal Text**
Papa en Marloes staan op het station.
Ze wachten op de trein.
Eerst hebben ze een kaartje gekocht.
Er stond een hele lange rij, dus dat duurde wel even.
Nu wachten ze tot de trein eraan komt.
Het is al vijf over drie, dus het duurt nog vier minuten.
Er staan nog veel meer mensen te wachten.
Marloes kijkt naar links, in de verte ziet ze de trein al aankomen.

**Oral Text**
Het is zaterdag.
Els heeft vrij.
Ze loopt door de stad.
Het is prachtig weer, de lucht is blauw.
Op straat ziet ze Bart op de fiets.
Hij wacht voor het rode licht.
Als Bart haar ziet, zwaait hij.
Els loopt weer verder.
Bij de bakker koopt ze brood, bij de slager koopt ze vlees.
Als het vijf uur is, gaat ze terug, zodat ze op tijd weer thuis is.
CHAPTER 5

INFLUENCE OF GENDER AND AGE ON THE NASALITY SEVERITY INDEX 2.0 IN DUTCH-SPEAKING FLEMISH CHILDREN AND ADULTS


Abstract

**Objective.** This study aimed to explore the influence of gender and age on the Nasality Severity Index 2.0 (NSI 2.0), an instrumental multiparametric index to determine hypernasality, in order to establish reference values for this new index.

**Methods.** Influence of gender and age on the NSI 2.0 was explored in 80 Dutch-speaking Flemish children (4-12y; 40 boys, 40 girls) and 60 Dutch-speaking Flemish adults (18-60y, 30 men, 30 women) without resonance disorders by determining its incorporated acoustic parameters: nasalance of the vowel /u/ and an oral text, determined by a Nasometer, and voice low tone to high tone ratio (VLHR) of the vowel /i/. The equation yields NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%) – (0.26 x nasalance oral text (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)).

**Results.** No effect of gender or age was found on the NSI 2.0 in children. However, significant differences were found for the NSI 2.0, nasalance of /u/ and an oral text between adult men and women. Additionally, an interaction effect between gender and age group was found for these parameters. Consequently, separate reference values for the NSI 2.0 in children, adult men and adult women were established.

**Conclusion.** The availability of reference values for the NSI 2.0 can be considered a next step in the implementation of this new, instrumental approach in the assessment of patients with hypernasality. Based on these reference scores, deviation of resonance in patients with resonance disorders can be defined. Further research can explore the possible influence of language on the index.

**Key words.** Nasality Severity Index; reference values; hypernasality; gender; age
Introduction

As resonance is a complex phenomenon, a single parameter approach may be insufficient to adequately diagnose resonance disorders (Dalston et al., 1993; Keuning et al., 2002). More specifically, single instrumental assessment techniques cannot always discriminate patients with resonance disorders from persons without resonance disorders (Karling et al., 1993; Van Lierde et al., 2007) or do not provide a degree of the resonance disorder which prevents the determination of severity and the evaluation of therapy effects (Cairns et al., 1996; Maier et al., 2008; Rah et al., 2001; Vijayalakshmi et al., 2007). Therefore, most speech-language pathologists rely on a combination of perceptual, acoustic and physiological techniques for the assessment of resonance disorders (Stelck et al., 2011). However, contradictory results can emerge when the outcomes of different assessment techniques are compared. This complicates decisions for treatment planning and clear communication to the patient and other clinicians. Acoustic analyses based on, for example, nasometry or spectral analyses, do not always correlate strongly with perceptual judgments (Keuning et al., 2002; Lewis et al., 2003; Nellis et al., 1992; Prado-Oliveira et al., 2015; Watterson et al., 1993) or are based on vowels only, which may limit their representativeness of spontaneous speech (Lee et al., 2009; Rah et al., 2001; Vijayalakshmi et al., 2007). Therefore, perceptual judgments remain the gold standard in the evaluation of speech characteristics. However, several variables, such as the experience of the listener (Brunnegard et al., 2012; Lewis et al., 2003), articulation errors (Garcia et al., 2014; Keuning et al., 2002), type of rating scale (Whitehill et al., 2002) and vowel content of the speech sample (Watterson et al., 2007), can influence listeners’ perception of speech which may limit the reliability and validity of perceptual judgments.

A possible solution to sidestep the limitations of single indirect instrumental assessment techniques is the combination of different variables into a multiparametric index. Following this, Van Lierde et al. (2007) took a first step in creating an instrumental and multiparametric protocol to assess resonance disorders by constructing a ‘Nasality Severity Index’ (NSI), based on a combination of acoustic and aerodynamic parameters. However, influence of personal and environmental variables was detected, resulting in large standard deviations even in a population without resonance disorders (Bettens et al., 2013). Therefore, Bettens et al. (2016) suggested an adaptation of the index resulting in the development of the NSI 2.0. Based on the data of different instrumental measurement techniques and optimal statistical discrimination between 35 children with perceived hypernasality and 50 children without resonance disorders, a weighted linear combination of three variables was established (see Bettens et al. (2016) for more information about the rationale behind and the derivation of the formula). More specifically, the index includes the nasalance scores of the vowel /u/ and an oral text (devoid of nasal consonants and originally developed by Van de Weijer and Slis (1991)) determined with a Nasometer II model 6450, and voice low tone to high tone ratio (VLHR) of the vowel /i/, using a cutoff frequency of 4.47*F0Hz (Lee et al., 2003; Lee et al., 2006). The equation of the NSI 2.0 yields NSI 2.0 = 13.20 – (0.0824 x
Influence of gender and age on the NSI 2.0

The mean NSI 2.0 score in children with perceived hypernasality was -6.8 (SD 5.14), whereas the mean NSI 2.0 score in children with normal resonance was +4.1 (SD 1.59). The index objectively discriminated children with perceived hypernasality from children without resonance disorders with a sensitivity of 92% and a specificity of 100%, using a cutoff score of zero. In contrast, an overall sensitivity of 81% and specificity of 93% was found by the authors (Bettens et al., 2016), based on the nasalance scores of an oral text alone, which confirmed that the NSI 2.0 clearly identifies patients and controls more accurately in comparison with a single parameter approach. Additionally, the validity of this new index was proven to be high, given that the application of the derived index on the data of an independent patient and control group resulted in 88% sensitivity and 89% specificity (Bettens et al., 2016).

One of the necessary conditions to implement this new index in daily clinical practice and to evaluate interventions properly is the availability of normative values derived from children and adults without resonance disorders. To establish these reference values, the possible influence of gender and age on the index needs to be explored. Two of the three variables included in the NSI 2.0 are obtained by a Nasometer II model 6450. This device, originally developed by Fletcher and Bishop (1973) and manufactured by Kay Pentax (NJ, Lincoln Park), is frequently used for the instrumental assessment of resonance, and is considered an indirect measure of nasality. In literature, contradictory results have been reported regarding the influence of gender and age on nasalance scores (see Bettens et al. (2013) for a recent overview). Regarding the parameters included in the NSI 2.0, Van Lierde et al. (2003) reported significantly higher nasalance scores for the vowel /u/ in Dutch-speaking Flemish women compared to men. Luyten et al. (2012) and Okalidou et al. (2011), on the other hand, reported no significant influence of gender on this parameter in children (Luyten et al., 2012) or in adults (Okalidou et al., 2011). Additionally, no significant effect of age was found on the nasalance score of this vowel (Luyten et al., 2012; Van Lierde et al., 2003). For the nasalance scores of an oral text, only Prathanee et al. (2003) found a significant difference between boys and girls. All other authors reported no significant influence of gender on the nasalance score for this parameter in children (Brunnegard & van Doorn, 2009; Luyten et al., 2012; Nichols, 1999; Rochet et al., 1998; Sweeney et al., 2004; Van de Weijer & Slis, 1991; Van der Heijden et al., 2011; Van Doorn & Purcell, 1998; Van Lierde et al., 2003), or in adults (Awan et al., 2015; Hirschberg et al., 2006; Nichols, 1999; Okalidou et al., 2011; Rochet et al., 1998; Seaver et al., 1991; Tachimura et al., 2000; Van de Weijer & Slis, 1991; Van Lierde et al., 2003). More controversy exists about the effect of age on the nasalance score of an oral text or oral sentences. Hirschberg et al. (2006) and Rochet et al. (1998) highlighted a significant difference between the scores of children and adults in which children showed lower nasalance values, while Nichols (1999), Van de Weijer and Slis (1991), and Van Lierde et al. (2003) found no significant differences between children and adults for this parameter. Considering adults only, Seaver et al. (1991) found a weak, positive correlation (r=0.311) between age and the nasalance value for an oral text. When only
children’s results are considered, no effect of age on the nasalance scores for an oral text has been reported (Brunnegard & van Doorn, 2009; Luyten et al., 2012; Van der Heijden et al., 2011; van Doorn & Purcell, 1998).

The third parameter included in the NSI 2.0 is ‘VLHR of the vowel /i/’ with a cutoff frequency of 4.47*F0Hz (Lee et al., 2003), determined by PRAAT software (Boersma & Weenink, 2014). This technique is based on the introduction of pole-zero pairs in the spectrum due to the coupling of the nasal to the oral tract (Lee et al., 2009). As a result of this coupling, an extra ‘nasal’ formant, the pole, appears in the neighborhood of the first formant, and the inclusion of the nasal sinuses as a resonance area is assumed to create anti-resonance, i.e. the zero (Feng & Castelli, 1996). A power ratio with a specific cutoff frequency between this pole and zero can therefore measure the amount of coupling of the nasal tract (Lee et al., 2009). Because formant frequencies change with age due to the growth of the vocal tract and resonance areas (Flipsen & Lee, 2012; Vorperian & Kent, 2007), a static cutoff frequency may be insufficient to properly detect energy changes in the spectrum of children. Therefore, a cutoff based on the fundamental frequency (F0) was preferred as F0 generally depends on body structure (Vorperian & Kent, 2007) and is therefore indirectly related to the formant frequencies. Following Lee et al. (2003), a cutoff frequency of 4.47*F0Hz was chosen by Bettens et al. (2016). Due to the use of a variable cutoff frequency based on the fundamental frequency and a relative rather than an absolute index, no influence of gender or age is hypothesized. However, no study yet investigated the possible influence of these variables.

The aim of the present study was to explore the influence of gender and age on the NSI 2.0 and its parameters in Dutch-speaking Flemish children and adults without resonance disorders. Depending on the results of this research question, reference values for the NSI 2.0 will be established for different gender and/or age groups or for the whole group of children and adults in total. Based on these reference scores, deviation of resonance in patients with resonance disorders can be defined. Moreover, it will be possible to compare the outcome of surgical interventions and speech therapy with these normative values. For example, after speech improving surgery, such as a palatal re-repair or velopharyngeal flap surgery, the remaining amount of deviation from the norm can be identified. Furthermore, the impact of adenotomy and/or tonsillectomy on nasal resonance can be verified by comparing the patient’s results with the normative values. Based on this information, the need for further intervention can be determined. Some influence of gender and age is hypothesized based on literature, in which women may show lower NSI 2.0 scores compared to men, and children may show higher NSI 2.0 scores compared to adults, resulting from differences in nasalance scores.

**Method**

This research study was approved by the institutional review and ethical board of the Ghent University Hospital (EC/2012/049).
**Participants**

Informed consents and questionnaires were sent to the parents of 937 children between 4 and 12 years old from four different primary schools in Flanders, the northern part of Belgium. A total of 275 signed informed consents and completed questionnaires were returned (response rate 29.3%). A total of 89 adult participants were recruited via friends, family or colleagues from the department of Speech, Language and Hearing Sciences at the Ghent University. According to the questionnaires, participants were selected for further analysis based on the following criteria: being aged between 4 and 12 years old (children) or between 18 and 60 years old (adults), being a native speaker of Dutch, living in Flanders, absence of moderate to severe hearing problems, allergy, neurological or velopharyngeal disorder, absence of a developmental delay, general disability, orthodontic treatment or oral surgical intervention. Ninety-two participants were excluded, more specifically due to not being a native speaker of Dutch (n=4), having a developmental delay (n=3), allergy (n=10), orthodontic treatment (n=9) or oral surgical intervention (n=66). To determine the speech performance of the selected participants, a 5-minute sample of conversational speech was collected and judged by a speech-language therapist (K.B.). If a resonance disorder, articulation error or voice problem was observed, the participant was excluded from the study. Twenty-one participants were judged with hyponasality due to a cold, eight were hypernasal, 12 suffered an articulation error and 24 were dysphonic. One-hundred twenty-two children and 85 adults were selected for further analysis, however, eight of the children were non-cooperative. A total of 114 children (40 boys, 74 girls) and 85 adults (30 men, 55 women) completed all assessments successfully. To create equal groups controlled for gender and age, 40 girls were matched in ages with the 40 boys, resulting in a group ranging from 4y1m to 11y10m years old (boys: mean 8.3y, SD 2.17; girls: mean 8.1y, SD 2.20); 30 women were matched in ages with the 30 men, resulting in a group ranging from 18y to 60y old (men: mean 34.5y, SD 13.05; women; mean 35.6y, SD 10.27).

**Procedure**

In order to determine the Nasality Severity Index 2.0, its parameters were defined using different instrumental assessment techniques. All assessments were performed in a quiet room at the child’s school or at the department of Speech, Language and Hearing Sciences at the Ghent University.

**Nasalance values.** To collect the nasalance values of the vowel /u/ and an oral text, a Nasometer (model II 6450 IBM PC version, Kay Pentax, NJ, Lincoln Park), developed by Fletcher and Bishop (1973), was applied. A sound separation plate, including two microphones to capture the nasal and oral signals, was placed between the upper lip and the nose of the participant. After calibration of the device as described by the instruction manual, the participant was asked to sustain the vowel /u/ and to read or repeat a standardized oral text at comfortable loudness and pitch. Nasalance scores were determined for both speech stimuli by taking the amount of nasal energy in relation to the amount of...
nasal-plus-oral energy and multiplying this ratio by 100. The oral passage, originally described by Van de Weijer and Slis (1991), represents most oral speech sounds in Dutch (Van de Weijer & Slis, 1991) and is based on the zoo passage in English (Fletcher, 1978).

**Acoustic measurements.** To determine VLHR (Lee et al., 2003; 2006), each participant was asked to sustain the vowel /i/ for at least 2s. To that end, a unidirectional microphone (Samson, C01U) was placed 10cm from the participant’s mouth. Recordings were made using free PRAAT software, version 5.3.78 (Boersma & Weenink, 2014), at a sampling frequency of 44100Hz. All signals were analyzed based on a 0.5-second fragment, extracted using a Hamming window. The first 0.5s of the sound sample was removed to eliminate the voice onset (Lee et al., 2003). Based on a cutoff frequency of 4.47*F0Hz, the spectrum was divided into a low frequency (65Hz to cutoff) and a high frequency (cutoff to 8000Hz) spectral region. After summation of the power of each frequency region, VLHR was determined as the power ratio of the low frequency to the high frequency spectrum in accordance with Lee et al. (2003; 2006). This analysis was completed using a PRAAT script to facilitate the proceeding.

**NSI.** The equation of the NSI 2.0 yields NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%) – (0.260 x nasalance oral text (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)). To define the NSI 2.0 for each participant, the results of the above-mentioned assessments were put into the formula.

**Statistical analysis.** IBM SPSS Statistics software version 22.0 (IBM Corp., Armonk, NY) was used for statistical analysis. Analysis of Covariance (ANCOVA) was applied to investigate the effect of gender on the NSI 2.0 and its parameters, with age as a covariate, in both the children’s and adults’ group. To verify a possible difference between children and adults, a two-way Analysis of Variance (ANOVA) was applied using gender and age group as fixed factors. As the parameters ‘nasalance of /u/’ and ‘nasalance of an oral text’ were not normally distributed, a logarithmic transformation was performed after which the assumption of normality was fulfilled. The assumption of homogeneity of the regression slopes was verified and accepted for the NSI 2.0 and all its parameters (p>0.05). A probability level less than 0.05 was considered to be significant. Additionally, effect sizes were defined based on a Pearson correlation coefficient (Cohen, 1992; Field, 2012).

**Results**

**Effect of gender on children’s and adults’ NSI 2.0 scores and its parameters**

Tables 5.1 and 5.2 summarize the results of the gender and age effect on the NSI 2.0 and its parameters for children and adults, respectively. No statistically significant gender effect on the NSI 2.0 was identified in children after controlling for the effect of age (p>0.05) (Figure 5.1a). However, in adults, a significant effect of gender was detected (p<0.01), in which women had significantly lower NSI 2.0 scores compared to men (Figure 5.1b). Additionally, the analysis of the different parameters of the NSI 2.0 revealed no statistically
significant influence of gender on the parameters 'nasalance of /u/', 'nasalance of an oral text' and 'VLHR of /i/' in children, and the parameter 'VLHR of /i/' in adults ($p>0.05$), when the covariate age was taken into account. However, for the parameters 'nasalance of /u/' and 'nasalance of an oral text', statistically significant higher scores were found for women compared to men ($p<0.01$), without an effect of age. Based on these findings, reference values for the NSI 2.0 and its parameters were established (Table 5.3).

**Effect of age on children’s and adults’ NSI 2.0 scores and its parameters**

Based on the results of the statistical analyses presented in Tables 5.1 and 5.2, the covariate age was not significantly related to the NSI 2.0 in children, nor in adults ($p>0.05$). Additionally, no significant influence of age was detected on its parameters, both in children and adults, when the data of children and adults were compared within their own age group (Table 5.1 and 5.2). However, when children’s and adults’ data were compared, a significant interaction effect between gender and age group was found for the NSI 2.0 (ANOVA: $F(1,136)=8.955$, $p=0.003$, $r=0.23$), and some of its parameters, more specifically, nasalance of /u/ (ANOVA: $F(1,136)=8.818$, $p=0.004$, $r=0.22$) and nasalance of an oral text (ANOVA: $F(1,136)=11.029$, $p=0.001$, $r=0.25$). As can be derived from Figure 5.1, adult men showed significantly higher NSI 2.0 scores compared to adult women (Post hoc Scheffé test: $p=0.014$), girls (Post hoc Scheffé test: $p=0.024$) and boys (Post hoc Scheffé test: $p=0.003$). Furthermore, significantly lower nasalance values for /u/ were observed in adult men compared to adult women (Post hoc Scheffé test: $p=0.003$), girls (Post hoc Scheffé test: $p=0.003$) and boys (Post hoc Scheffé test: $p=0.002$). Additionally, significantly higher nasalance values for an oral text were observed in adult women compared to girls (Post hoc Scheffé test: $p<0.001$), and borderline significant differences for this parameter were observed in adult women compared to boys (Post hoc Scheffé test: $p=0.060$) and adult men (Post hoc Scheffé test: $p=0.063$). Regarding VLHR of /i/, no interaction was found between gender and age group (ANOVA: $F(1,136)=0.130$, $p=0.719$, $r=not applicable due to negative variance estimate$). Therefore, the main effect of gender and age group on this parameter was investigated, resulting in an effect of age group (ANOVA: $F(1,137)=24.743$, $p<0.001$, $r=0.38$) without an effect of gender (ANOVA: $F(1,137)=0.594$, $p=0.442$, $r=not applicable due to negative variance estimate$), in which adults showed lower VLHR scores compared to children.
### Table 5.1 Influence of gender and age on the Nasality Severity Index 2.0 (NSI 2.0) and its parameters in children.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Boys (n=40)</th>
<th>Girls (n=40)</th>
<th>Effect of gender (ANCOVA)</th>
<th>Effect of age (ANCOVA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>95% CI</td>
<td>F</td>
</tr>
<tr>
<td>Nasalance /u/ (%)</td>
<td>11.4†</td>
<td>0.43</td>
<td>9.0-14.4‡</td>
<td>0.039</td>
</tr>
<tr>
<td>Nasalance oral text (%)</td>
<td>11.0†</td>
<td>0.31</td>
<td>9.6-12.8‡</td>
<td>3.818</td>
</tr>
<tr>
<td>VLHR /i/ 4.47*F0Hz (dB)</td>
<td>21.65</td>
<td>5.15</td>
<td>20.00-23.30</td>
<td>0.097</td>
</tr>
<tr>
<td>NSI 2.0</td>
<td>3.9</td>
<td>1.67</td>
<td>3.4-4.5</td>
<td>0.653</td>
</tr>
</tbody>
</table>

*Significant difference, p<0.05; †Geometric mean resulting from logarithmic transformation and antilog; ‡Based on Cox method (Zhou et al., 1997); F=F-ratio from ANCOVA, df: degrees of freedom, r: Pearson correlation coefficient - effect size.

### Table 5.2 Influence of gender and age on the Nasality Severity Index 2.0 (NSI 2.0) and its parameters in adults.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Men (n=30)</th>
<th>Women (n=30)</th>
<th>Effect of gender (ANCOVA)</th>
<th>Effect of age (ANCOVA)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>95% CI</td>
<td>F</td>
</tr>
<tr>
<td>Nasalance /u/ (%)</td>
<td>6.9†</td>
<td>0.61</td>
<td>4.5-10.5‡</td>
<td>7.437</td>
</tr>
<tr>
<td>Nasalance oral text (%)</td>
<td>10.9†</td>
<td>0.26</td>
<td>9.6-12.4‡</td>
<td>11.082</td>
</tr>
<tr>
<td>VLHR /i/ 4.47*F0Hz (dB)</td>
<td>17.05</td>
<td>5.58</td>
<td>15.00-19.13</td>
<td>0.653</td>
</tr>
<tr>
<td>NSI 2.0</td>
<td>5.5</td>
<td>1.57</td>
<td>4.9-6.1</td>
<td>10.756</td>
</tr>
</tbody>
</table>

*Significant difference, p<0.05; †Geometric mean resulting from logarithmic transformation and antilog; ‡Based on Cox method (Zhou et al., 1997); F=F-ratio from ANCOVA, df: degrees of freedom, r: Pearson correlation coefficient - effect size.
Table 5.3 Reference values for the Nasality Severity Index (NSI 2.0) and its parameters for the total group of children, and adult men and women separately.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Children (n=80)</th>
<th>Men (n=30)</th>
<th>Women (n=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>95% PI</td>
</tr>
<tr>
<td>Nasalance /u/ (%)</td>
<td>11.3†</td>
<td>0.46</td>
<td>4.5-28.1</td>
</tr>
<tr>
<td>Nasalance oral text (%)</td>
<td>10.4†</td>
<td>0.59</td>
<td>5.7-18.8</td>
</tr>
<tr>
<td>VLHR /i/ 4.47*F0Hz (dB)</td>
<td>21.84</td>
<td>5.06</td>
<td>11.86-31.83</td>
</tr>
<tr>
<td>NSI 2.0</td>
<td>4.1</td>
<td>1.63</td>
<td>0.9-7.3</td>
</tr>
</tbody>
</table>

†Geometric mean resulting from logarithmic transformation and antilog
Figure 5.1 NSI 2.0 in function of gender and age in normal-Dutch-speaking Flemish children and adults. 95% prediction intervals are provided. ANCOVA revealed no significant influence of gender and age on NSI 2.0 scores in children. In adults, men showed significantly higher NSI 2.0 scores compared to women, without an effect of age. Comparing children's and adults' scores, ANOVA revealed a statistically significant interaction effect between gender and age on the NSI 2.0 scores ($p>0.05$), in which adult men showed higher NSI 2.0 scores compared to adult women, girls and boys.

a. NSI 2.0 in function of gender and age in children (boys, n=40; girls, n=40).
b. NSI 2.0 in function of gender and age in adults (men, n=30; women, n=30).

Discussion

This study aimed to explore the influence of gender and age on the NSI 2.0 in order to create reference values for this new, instrumental and multiparametric index to determine hypernasality, in Dutch-speaking Flemish children and adults. No statistically significant influence of gender and age was detected on the NSI 2.0 and its parameters in children, which is in accordance with literature. More specifically, no influence of gender and age was reported in children for nasalance values of the vowel /u/ (Luyten et al., 2012) and nasalance values of an oral text or oral sentences (Brunnegard & van Doorn, 2009; Karakoc et al., 2013; Luyten et al., 2012; Nichols, 1999; Park et al., 2014; Prathanee et al., 2003; Sweeney et al., 2004; Van de Weijer & Slis, 1991; Van der Heijden et al., 2011; van Doorn & Purcell, 1998; Van Lierde et al., 2003). However, significantly lower NSI 2.0 scores were observed in women compared to men, indicating the presence of more nasal resonance in women. This resulted from the higher nasalance values of the vowel /u/ and an oral text in women. Regarding nasalance of /u/, Van Lierde et al. (2003) also found significantly higher scores for this parameter in women compared to men. However, no difference between men and women has yet been described in literature for the nasalance value of an oral text or oral sentences.
Influence of gender and age on the NSI 2.0

(Hirschberg et al., 2006; Nichols, 1999; Okalidou et al., 2011; Rochet et al., 1998; Seaver et al., 1991; Tachimura et al., 2000; Van de Weijer & Slis, 1991; Van Lierde et al., 2003), although higher nasalance values have been observed in women for stimuli including nasal consonants (i.e. an oronasal text (Rochet et al., 1998; Van Lierde et al., 2003) and/or nasal text (Seaver et al., 1991; Van Lierde et al., 2003)). Possible explanations may be the underlying anatomical and physiological differences related to the velopharyngeal closure between men and women (Kuehn, 1976; Seaver et al., 1991), or an existing mismatch in sensitivity of the two microphones of the Nasometer for certain fundamental frequencies (Zajac et al., 1996). However, despite the medium effect sizes (Cohen, 1992), the reported differences are small and could also be due to intrasubject variability (Brunnegard & van Doorn, 2009). When looking at the influence of age in adults, no influence was detected on the NSI 2.0 scores and its parameters, which is comparable with Seaver et al. (1991), although they reported a weak correlation between age and nasalance scores of an oral text.

When the data of children and adults were compared, a significant interaction between gender and age group was found for the NSI 2.0 scores, in which adult men showed higher NSI 2.0 scores compared to adult women and children, indicating less nasal resonance in the speech of adult men. This resulted from lower nasalance values for /u/ and lower VLHR scores for /i/ observed in men. Only Van Lierde et al. (2003) yet reported on the comparison of nasalance scores on vowels between children and adults. They found no significant age effect on the nasalance scores of the vowel /u/, which is in contrast with the findings of the current study. Furthermore, higher nasalance values for an oral text were observed in women compared to adult men and children in the current study. In literature, also higher values have been reported for nasalance values of an oral text or oral sentences in adults compared to children, however without an interaction of gender (Hirschberg et al., 2006; Rochet et al., 1998). The found differences could be due to underlying anatomical and physiological differences related to the velopharyngeal closure (Hirschberg et al., 2006; Rochet et al., 1998), such as atrophy of lymphoid structures with increasing age (Becker et al., 2009) and growth of the oropharynx (Taylor et al., 1996), whether or not in combination with an existing mismatch in sensitivity of the two microphones of the Nasometer for certain fundamental frequencies (Zajac et al., 1996) or intrasubject variability (Brunnegard & van Doorn, 2009), as the differences were small. Considering the VLHR of the vowel /i/, lower scores were observed in adults compared to children, without an effect of gender, indicating the presence of less nasal resonance in adults on this specific vowel. As no difference was observed for the VLHR of /i/ between adult men and women, but lower nasalance values for /u/ were found in men, this inconsistency may confirm the possible influence of the fundamental frequency on nasalance values, as the difference in fundamental frequency was controlled for by the variable cutoff score based on F0 to determine VLHR of /i/.

Although the found differences were statistically significant, the differences were small and only low to medium effect sizes were found. Therefore, the clinical relevance of the found differences can probably be considered as limited. Nevertheless, based on these study
outcomes, separate references values for the NSI 2.0 and its parameters were established for children, adult men and adult women (Table 5.3, p. 81).

All participants’ scores exceeded zero (range +0.1 to +7.3 in children, range +0.6 to +8.7 in adults), which is comparable with the results of the control group in the study by Bettens et al. (2016) in which all children without resonance disorders presented with NSI 2.0 scores above zero. Additionally, a smaller standard deviation of the mean (SD 1.63) was observed in children compared with the standard deviations of the original NSI developed by Van Lierde et al. (2007) (SD 13.8 in boys and SD 18.8 in girls (Bettens et al., 2013)). Bettens et al. (2013) mentioned that a large spread of NSI values in children without resonance disorders indicates too large inter-individual differences to consider the original NSI as a reliable instrument to assess resonance. Due to the replacement of the previously included parameters ‘maximum duration time of /s/’ and ‘mirror-fogging test by Glatzel of /a/’ by the new parameter ‘VLHR of /i/ with a cutoff frequency of 4.47*F0Hz’, the influence of personal and environmental variables was reduced, resulting in a smaller standard deviation of the mean in the present study.

Several studies reported on the influence of language (Van Lierde et al., 2001) and dialect (Awan et al., 2015; Mayo et al., 1996; Nichols, 1999; Seaver et al., 1991) on nasalance values. As no influence of dialect has been reported on nasalance values in Dutch (D’haeseleer et al., 2015), the reference values reported in the current study can be applied among patients with different Dutch dialects. However, based on the study by Van Lierde et al. (2001), nasalance values for an oral text in Dutch-speaking Flemish adults differ significantly from those of Canadian English (Kavanagh et al., 1994), American English (Seaver et al., 1991) and Spanish (Anderson, 1996). Between Dutch spoken in Belgium and in the Netherlands, no difference was found for this parameter (Van de Weijer & Slis, 1991). Hence, further research can explore the need for reference values for the NSI 2.0 in different languages.

Conclusion

The availability of reference values for the NSI 2.0 can be considered as a next step in the implementation of this new, instrumental approach in the assessment and treatment planning of patients presenting with hypernasality. Based on these reference scores, deviation of resonance in patients with resonance disorders can be defined. As differences were found in NSI 2.0 scores and its parameters between children and adults, and between adult men and women, separate references values were established for children, adult men and adult women. Although no straightforward explanations could be provided for the observed differences, the differences were small, which probably limits their clinical relevance. As previous studies highlighted differences between nasalance values in different languages, further research can explore the need for reference values for the NSI 2.0 in various languages.
Reducing the results of different indirect assessment techniques into one NSI 2.0 value sidesteps the limitations of a single parameter approach, but preserves the distinctness of the interpretation of a single value, especially for the patient, his/her relatives and clinicians other than speech-language pathologists. Nevertheless, the inclusion of a perceptual evaluation and physiological assessment remains absolutely necessary to confirm the results of acoustic measurements and to visualize the function of the velopharyngeal mechanism.

**Acknowledgements**

The authors gratefully acknowledge all children, their parents and all adult participants for their contribution, as well as the directors and teachers of the participating schools.

**Declaration of interest.** The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.
References


CHAPTER 6

SHORT-TERM AND LONG-TERM TEST-RETEST RELIABILITY OF THE NASALITY SEVERITY INDEX 2.0


Abstract

Objective. The Nasality Severity Index 2.0 (NSI 2.0) forms a new, multiparametric approach in the assessment of hypernasality. To enable clinical implementation of this index, the short- and long-term test-retest reliability of this index was explored.

Methods. In 40 normal-speaking adults (mean age 32y, SD 11, 18-56y) and 29 normal-speaking children (mean age 8y, SD 2, 4-12y), the acoustic parameters included in the NSI 2.0 (i.e. nasalance of the vowel /u/ and an oral text, and the voice low tone to high tone ratio (VLHR) of the vowel /i/) were obtained twice at the same test moment and during a second assessment two weeks later. After determination of the NSI 2.0, a comprehensive set of statistical measures was applied to determine its reliability.

Results. Long-term variability of the NSI 2.0 and its parameters was slightly higher compared to its short-term variability, both in adults and in children. Overall, a difference of 2.82 for adults and 2.68 for children between the results of two consecutive measurements can be interpreted as a genuine change. With an ICC of 0.84 in adults and 0.77 in children, the NSI 2.0 additionally shows an excellent relative consistency. No statistically significant difference was withheld in the reliability of test-retest measurements between adults and children.

Conclusion. Reliable test-retest measurements of the NSI 2.0 can be performed. Consequently, the NSI 2.0 can be applied in clinical practice, in which successive NSI 2.0 scores can be reliably compared and interpreted.

Key words. Nasality Severity Index; hypernasality; reliability

Learning outcomes. The reader will be able to describe and discuss the short-term and long-term test-retest reliability of the Nasality Severity Index 2.0, a new multiparametric approach to assess hypernasality, and its parameters. Based on this information, the NSI 2.0 can be applied in clinical practice, in which successive NSI 2.0 scores, e.g. before and after surgery, can be compared and interpreted.
Introduction

To assess and diagnose hypernasality, speech-language pathologists as well as other clinicians mostly rely on a combination of perceptual and instrumental measurements. A perceptual assessment based on spontaneous speech, automatic speech and reading or repeating sentences and words remains the “gold” standard to determine resonance disorders. However, perceptual measurements are subjective and therefore can be influenced by vocal quality (Kataoka et al., 2001) and articulation errors of the patient (Bzoch, 1997) or by experience of the examiner (Lewis et al., 2003). To support the perceptual analysis, several instrumental measurements are available to determine the presence and degree of resonance disorders (Bettens et al., 2014). However, contradictory results can emerge when the outcomes of different assessment techniques are compared. Acoustic analyses based on, for example, nasometry or spectral analyses do not always strongly correlate with perceptual judgments (Keuning et al., 2002; Lewis et al., 2003; Nellis et al., 1992; Prado-Oliveira et al., 2015; Wattersen et al., 1993) or are based on vowels only, which may limit their representativeness of spontaneous speech (Lee et al., 2009; Rah et al., 2001; Vijayalakshmi et al., 2007). Therefore, a combination of complementary test results into a multiparametric index can form a solution. In a pilot study, Van Lierde et al. (2007) developed a ‘Nasality Severity Index’ (NSI) based on a combination of five parameters, more specifically the nasalance value of the vowel /a/, an oral and oronasal text derived by the Nasometer (model 6200); the maximum duration time (MDT) of /s/; and the mirror-fogging test by Glatzel of the vowel /a/. The equation yielded $\text{NSI} = -60.69 - (3.24 \times \text{nasalance oral text (\%)} - (13.39 \times \text{Glatzel value /a/}) + (0.244 \times \text{MDT /s/ (s)}) - (0.558 \times \text{nasalance /a/ (\%)}) + (3.38 \times \text{nasalance oronasal text (\%)})$. However, influence of personal and environmental variables due to the inclusion of the MDT of /s/ and the use of the mirror-fogging test by Glatzel (Foy, 1910) was detected (Bettens et al., 2013). Therefore, Bettens et al. (2016) proposed an adaptation of the NSI based on the data of different instrumental measurement techniques and the optimal statistical discrimination between 50 children without resonance disorders and 35 children with hypernasality, in a stepwise statistical approach, with sensitivity and specificity as the serving criteria. A weighted linear combination of three variables was established, more specifically the nasalance scores of the vowel /u/ and an oral text obtained with the Nasometer (model II 6450) and the voice low tone to high tone ratio (VLHR) of the vowel /i/ with a cutoff frequency of 4.47*F0Hz (originally described by Lee et al. (2003)) (see Bettens et al. (2016) for more information about the rationale behind and the derivation of the formula). The formula of the adapted NSI yields $\text{NSI 2.0} = 13.20 - (0.0824 \times \text{nasalance /u/ (\%)}) - (0.260 \times \text{nasalance oral text (\%)}) - (0.242 \times \text{VLHR /i/ 4.47*F0Hz (dB)})$. The mean NSI 2.0 value of patients with perceived hypernasality was -6.8 (SD 5.14), whereas the mean NSI 2.0 value of the control children with normal resonance was +4.1 (SD 1.59). With a cutoff score of zero, the NSI 2.0 discriminated patients with hypernasality from persons without hypernasality with a sensitivity of 92% and a specificity of 100%, in which patients with perceived hypernasality had scores below zero. The validity of this new index was proven to be high by application of the parameter results of an independent patient and
control group on the derived formula (sensitivity 88%, specificity 89%), in which all patients were perceptually judged with hypernasality and all control children with normal resonance.

However, before the NSI 2.0 can be implemented in daily clinical practice, the reliability of this new index has to be verified. According to literature, several sources can affect the stability of instrumental measurements (Lewis et al., 2008). More specifically, instrumental variance (e.g. microphone and sound cart characteristics, machine model), test procedure (e.g. distance from the microphone), subject performance (e.g. physiological factors, nasal patency) and the environment (e.g. air moisture and temperature) can influence the reliability of assessment techniques. Similarly, the components of the NSI 2.0 are susceptible to these sources of variation.

Two of the three parameters included in the index are obtained by a Nasometer. This device, originally developed by Fletcher and Bishop (1970) and manufactured by KayPentax (KayPentax, NJ, Lincoln Park), determines the amount of nasal resonance based on an acoustic analysis of both a nasal and oral signal, and is considered an indirect measure of nasality. The signals are obtained by two microphones divided by a sound separation plate which is positioned between the nose and the upper lip of the participant. After filtering the signals using a band pass filter with a center frequency of 500Hz and a bandwidth of 300Hz, the ratio of the nasal signal to the (nasal + oral) signal, multiplied by 100, yields the nasalance score in a percentage. Several authors state that, although based on similar acoustic analyses of nasal and oral signals, nasalance scores of different instruments, such as the Nasometer, NasalView and OroNasal System, are not interchangeable (Awan et al., 2011; Bressmann, 2005; Bressmann et al., 2006; Lewis & Watterson, 2003). Additionally, scores obtained with different models of the same instrument can also vary significantly (Awan et al., 2011; Awan & Virani, 2013; de Boer & Bressmann, 2014; Watterson & Lewis, 2006). Even results determined with different devices of the same model may differ due to the characteristics of the nasal and oral microphone (Zajac et al., 1996). When the same device is used, replacement of the headgear can introduce a second source of variability (Watterson et al., 2005; Watterson & Lewis, 2006), although Lewis et al. (2008) and Kavanagh et al. (1994) found only small differences between the condition of no change of the headgear and headgear change between two successive measurements. Next to instrumental and procedure variation, personal variation also has an influence on the reliability of the test results. Extensive research focused on between-subject variability, more specifically the influence of age (Brunnegard & van Doorn, 2009; Luyten et al., 2012; Prathanee et al., 2003; Van der Heijden et al., 2011; van Doorn & Purcell, 1998), gender (Abou-Elsaad et al., 2012; Brunnegard & van Doorn, 2009; Karakoc et al., 2013; Luyten et al., 2012; Nichols, 1999; Park et al., 2014; Prathanee et al., 2003; Sweeney et al., 2004; Van de Weijer & Slis, 1991; Van der Heijden et al., 2011; van Doorn & Purcell, 1998; Van Lierde et al., 2003) and dialect (Awan et al., 2015; D’haeseleeer et al., 2015; Kavanagh et al., 1994; Mayo et al., 1996; Nichols, 1999; Rochet et al., 1998; Seaver et al., 1991). This resulted in normative values for nasalance scores in different languages.
Another personal inconsistency arises from intra-subject variability possibly due to the variation in physiological factors such as small changes in nasal patency (de Boer & Bressmann, 2014; Lewis et al., 2008; van Doorn & Purcell, 1998). Due to this variability, nasalance scores of an oral text or oral sentences differ from 4 to 6 percentage points in 95% of the recordings of participants with normal speech in the ‘no change of the headgear’ condition using a Nasometer (Lewis et al., 2008; Sweeney et al., 2004; van Doorn & Purcell, 1998; Watterson et al., 2005; Watterson et al., 2008) and from 5 to 9 percentage points in 95% of the recordings in the ‘change of the headgear’ condition between sessions (de Boer & Bressmann, 2014; Lewis & Watterson, 2003; Lewis et al., 2008; van Doorn & Purcell, 1998; Watterson et al., 2005; Whitehill, 2001). These studies are based on data obtained from adults or children. However, to the best of our knowledge, no study has yet investigated whether nasalance can be examined as reliable in children as in adults.

A third parameter included in the NSI 2.0 is VLHR of the vowel /i/ with a cutoff frequency of 4.47*F0Hz (Lee et al., 2003), determined by PRAAT software (Boersma & Weenink, 2014). After determination of the power spectrum of the sound wave by a Fast Fourier Transformation, the spectrum of the sound sample is divided into a low frequency band (LFB) and a high frequency band (HFB) using a specific cutoff frequency derived from the fundamental frequency, 4.47*F0Hz. To quantify LFB, the power of each frequency band ranging from 65Hz to the cutoff frequency is summed; the power of HFB is calculated as the summation of the power ranging from the cutoff frequency up to 8000Hz in accordance with the protocol of Lee et al. (2003; 2006). VLHR is defined as the power ratio of LFB vs. HFB, expressed in decibels. Lee et al. (2003) reported no significant correlation between sound intensity and VLHR, due to the use of a relative rather than an absolute index. Division of LFB to HFB also eliminates the possible influence of different sound recording conditions such as characteristics of the microphone, sound card of the computer and distance from the microphone. Therefore, variability due to equipment and test procedure may be limited. However, between-subjects and subject performance variability can still influence the stability of the test results for this parameter.

To enable the implementation of the NSI 2.0 in the diagnosis of hypernasality and evaluation of surgical treatment, such as palatal re-repair or speech therapy, the aim of this study was to explore both, short- and long-term test-retest reliability of the NSI 2.0 and its parameters in adults and children without resonance disorders. Furthermore, the possible difference in test-retest reliability between adults and children was verified in both conditions. Based on literature, long-term variability of the NSI 2.0 and its parameters is hypothesized to be larger compared to short-term variability. Additionally, variability of the NSI 2.0 and its parameters is expected to be larger in children than in adults.
Method

Participants

Forty-one adults between 18 and 56 years old (mean age 32y, SD 11), 29 women and 12 men, and 29 children between 4 and 12 years old (mean age 8y, SD 2), 16 girls and 13 boys, were included in this study. All adult participants were recruited via friends, family or colleagues from the department of Speech, Language and Hearing Sciences at the Ghent University. Children were recruited via their youth movement or primary school. According to a short questionnaire (orally completed by the adults, in written by the legal representative of the children), participants were selected based on the following criteria: being a native speaker of Dutch, living in Flanders (the northern part of Belgium), absence of moderate to severe hearing problems, neurological or velopharyngeal problem, absence of a developmental delay, general disability, orthodontic treatment or oral surgical intervention. None of the participants had a history of resonance disorders, nor reported having a cold or allergy outburst at the moment of testing. Informed consent was signed by all participants or their legal representative (ethical approval by the IRB). At the beginning of each assessment, the speech production of each participant was judged during a conversation of approximately 5 minutes by a speech-language pathologist (K.B.) with special attention to resonance. If a resonance disorder was observed (hypernasality, hyponasality, cul-de-sac resonance or mixed nasality), the participant was excluded from the study. One woman was judged with hypernasality and therefore excluded, resulting in an adult study group of 28 women and 12 men (mean age 32y, SD 11).

Instrumental assessment

To determine the parameters ‘nasalance score of /u/’ and ‘nasalance score of an oral text’, a Nasometer II model 6450 (KayPentax, NJ, Lincoln Park) was used. At the beginning of each test session, the device was calibrated in a quiet room following the instructions of the manufacturer’s manual. After the sound separation plate was placed between the nose and the upper lip, the participant was asked to sustain the vowel /u/ four times for two seconds during one recording according to the example of the investigator. The mean nasalance score of the vowel /u/ was derived from this recording using the software program of the Nasometer. Furthermore, each participant was asked to read an oral text with comfortable loudness and pitch. This oral reading passage, originally developed by Van de Weijer and Slis (1991), exclusively consists of oral speech sounds and is comparable with the zoo passage in English developed by Fletcher (1978). If a reading error was made, the participant was asked to read the passage again.

To obtain VLHR of /i/, each participant was asked to sustain the vowel /i/ for at least 2 seconds in front of a unidirectional microphone (Samson, C01U), placed at 10cm from the participant’s mouth. Recordings were made and analyzed using free PRAAT software, version 5.3.78 (Boersma & Weenink, 2014), at a sampling frequency of 44100Hz. After recording, a
0.5s fragment was selected using a Hamming window. The first 0.5s of the sound sample was removed to eliminate the voice onset (Lee et al., 2003). A cutoff frequency of 4.47*F0Hz was used to divide the spectrum into a low frequency (65Hz to cutoff) and high frequency (cutoff to 8000Hz) spectral region after Fast Fourier Transformation following Lee et al. (2003). After summation of the power of each frequency region, VLHR was determined as the power ratio of the low frequency to the high frequency spectrum, expressed in dB in accordance with Lee et al. (2003; 2006; 2009). The analysis was completed using a PRAAT script to facilitate the proceeding.

After determination of all three parameters, NSI 2.0 scores were derived for each participant at each test moment. The NSI 2.0 yields NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%) – (0.260 x nasalance oral text (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)) (Bettens et al., 2016).

**Measurements**

To determine the short-term reliability of the NSI 2.0 and its parameters, all participants were tested twice by the same investigator during the same test session. Between both measurements, the headgear of the Nasometer was removed and replaced to represent the clinical practice of the test procedure. To verify the long-term reliability of the index, 19 adults (13 women, 6 men; mean age 29y, SD 5) and 11 children (7 girls, 4 boys; mean age 7y, SD 2) were additionally assessed by the same investigator at a second test moment with a mean interval of 14 days (SD 4, range 7-28 days) between the first and second assessment. The drop-out of 39 participants was due to organizational reasons, e.g. impossibility of the participant to visit the department two times, and time limitations.

**Statistical procedure**

To determine the test-retest reliability of the NSI 2.0 and its parameters, a comprehensive set of statistical measures was applied using IBM SPSS Statistics software version 22.0 (IBM Corp., Armonk, NY). Before proceeding, assumptions of normality and absence of heteroscedasticity (i.e. amount of random error that increases when the measured values increase) were verified for all parameters and the NSI 2.0 for each test moment. To determine heteroscedasticity, Bland-Altman plots were drawn using the absolute differences and individual means between test and retest moments (Atkinson & Nevill, 1998; Bland & Altman, 1986). Subsequently, the Pearson correlation coefficient was derived. Heteroscedasticity was present when a statistically significant correlation was found (Pearson correlation coefficient: p<0.05). If heteroscedasticity was detected, a natural logarithmic transformation was performed on the data to fulfill the assumption of no heteroscedasticity.

The measurement error between two test moments consists of both systematic and random error. As systematic bias can be due to general learning, fatigue, a training effect or motivation, detection of this systematic error should induce an adaptation of the measurement protocol (Atkinson & Nevill, 1998). Therefore, repeated measurement analysis
of variance (ANOVA) was performed to verify possible systematic changes. Additionally, means and standard deviations of the differences between test and retest moments were determined. To investigate relative consistency, i.e. consistency of the position of an individual’s data in the group relative to others, the intraclass correlation coefficient (ICC) was determined using a 2-way random model with absolute agreement and single measures (Shrout & Fleiss, 1979; Weir, 2005). Nevertheless, the heterogeneity of the participants, represented by the between-subjects variability, has to be taken into account when interpreting the ICC. Data analysis of a homogeneous group, as in the assessment of normal participants, can induce lower ICC values compared to a more heterogeneous group despite similar levels of agreement (Costa-Santos et al., 2011). Therefore, the standard error of measurement (SEM), as an absolute index of reliability, was explored. The SEM is estimated by taking the square root of the mean square error term from the ANOVA (SEM=√MSE). This calculation was opted for to exclude the influence of the range of measured values (Atkinson & Nevill, 1998). Based on this statistic, the minimal detectable difference (MDD) was derived. To indicate the 95% confidence to detect a real difference between the true scores the equation MDD = SEM x 1.96 x √2 was used when the assumption of no heteroscedasticity was fulfilled (Beckerman et al., 2001; Weir, 2005). If heteroscedasticity was detected, the MDD was calculated using the formula SEM^{1.96} after the SEM was antilogged (Atkinson & Nevill, 1998; Bland & Altman, 1996). The MDD can be applied to describe the limits which should include the participant’s true value for 95% of the observations. If a second observation exceeds the 95% interval based on the first observation, the second observation may be considered as a result of a real difference, not due to personal variation or a measurement error. To calculate the 95% confidence interval (95% CI), the MDD has to be subtracted and added to the participant’s value when heteroscedasticity is absent. For heteroscedastic data, the participant’s value has to be divided and multiplied by the MDD because of the logarithmic scale (Bland & Altman, 1996).

To verify the possible difference in test-retest reliability between adults and children, an independent Student t-test was applied on the relative differences of the NSI 2.0 and its parameters in both, the short-term and long-term condition. A probability level less than 0.05 was considered to be significant.

**Results**

Normality tests (Kolmogorov-Smirnov test, Q-Q plots, and boxplots) revealed a normal distribution of all parameters and NSI 2.0 scores in both the adults’ and children’s group. No heteroscedasticity was withheld as the Pearson correlation coefficients of the Bland-Altman plots were not statistically significant (p>0.05), except for the parameter ‘nasalance of an oral text’ in the short-term condition of the children’s group (r=0.64, p<0.001) and the parameter ‘nasalance of /u/’ in the long-term condition of the adults’ group (r=0.72, p<0.001,). Therefore, a logarithmic transformation was performed on these
parameters after which the assumption of no heteroscedasticity was fulfilled ($r=0.32$, $p=0.09$; $r=-0.30$, $p=0.22$, resp.).

Table 6.1 outlines the results of the statistical measures to explore the test-retest reliability of the NSI 2.0 and its parameters, including the mean differences ($\text{mean}_{\text{diff}}$), standard deviations of the differences ($\text{SD}_{\text{diff}}$), the intraclass correlation coefficient (ICC), the standard error of measurement (SEM), and the minimal detectable difference to determine a confidence interval of 95% ($\text{MDD}_{95}$). No significant difference was observed in the short-term and long-term measurements in both groups (repeated measurements ANOVA, $p>0.05$), indicating the absence of a systematic error.

Comparing the mean differences and their SDs of the short- and long-term reliability results in similar mean differences, but higher standard deviations in the long-term condition for all parameters and the NSI 2.0 in both the adults’ and children’s group. Excellent agreement (ICC ≥ 0.75, Cicchetti (1994)) was found for the NSI 2.0 and its parameters based on the ICCs in the short-term condition, except for the parameter ‘VLHR of /i/’ in the children’s group, for which good agreement was found. Moreover, excellent agreement was also observed in the long-term condition for the NSI 2.0 and the parameters ‘nasalance of /u/’ and ‘nasalance of an oral text’ in both the adults’ and children’s group, whereas good agreement was obtained for ‘VLHR of /i/’. Comparison of the ICCs in the same condition reveals comparable values for the NSI 2.0 and the parameters ‘nasalance of /u/’ and ‘nasalance of an oral text’, in contrast with the somewhat lower values obtained for the parameter ‘VLHR of /i/’ in both the short- and long-term condition. Nevertheless, the ICC of the VLHR remains good to excellent for both the adults’ and children’s group in the short-term (ICC=0.85; ICC=0.74, resp.) and the long-term condition (ICC=0.73; ICC=0.63, resp.).

As the SEM and MDD depend on the unit of each parameter, mutual evaluation of the parameters is impossible. Again, somewhat higher values were detected in the long-term condition compared with the short-term condition for these statistical parameters, resulting in larger 95% CIs and confirming the larger variability for all parameters and the NSI 2.0 in the long-term condition for both the adults’ and children’s group.

No statistically significant differences were found between adults and children regarding the mean differences of the NSI 2.0 and its parameters in the short-term and long-term condition (independent Student t-test: $p>0.05$). Furthermore, comparable ICC values were observed, although the ICC for the parameters ‘nasalance of /u/’ in the short-term condition and ‘VLHR of /i/’ in the short- and long-term condition were a bit lower in the children’s group.

To enable a comparison with the literature, cumulative frequencies of the nasalance scores of the vowel /u/ and the oral text for both the short- and long-term reliability in adults and children are presented in Table 6.2. Taken the cutoffs at which 90% of the nasalance score differences are included, overall nasalance difference scores in adults and children are
comparable, which is in accordance with the non-significant nasalance difference scores between adults and children mentioned above. Comparing the adults’ results for the parameter ‘nasalance of /u/’ in the short-term and long-term condition reveals a larger variability in the long-term condition. For the parameter ‘nasalance of an oral text’ on the other hand, cumulative frequencies for both conditions are quite similar, confirming the small differences between the MDDs of the oral text in both conditions.

**Discussion**

To enable the implementation of the NSI 2.0 in the diagnosis of hypernasality and evaluation of intervention, this study aimed to verify the short- and long-term reliability of the NSI 2.0 and its parameters in adults and children without resonance disorders. Additionally, the possible difference in test-retest reliability between adults and children was explored. NSI 2.0 scores and its parameters were determined twice within one session and between two sessions separated by approximately two weeks. No statistically significant differences between adults and children were found for the difference scores of the NSI 2.0 and its parameters. Furthermore, the mean differences and their standard deviations in both groups were slightly higher for the long-term condition compared to the short-term condition. Additionally, the differences between both measurements in the short-term condition and those in the long-term condition were not statistically significant, resulting in the absence of a systematic error in either condition. As mentioned before, the measurement error between two test moments consists of both systematic and random error. The absence of a systematic error suggests no influence of a training effect, fatigue or motivation. Additionally, this emphasizes the accuracy of the applied test procedure. Differences between two measurements could be explained by a random error caused by repositioning the Nasometer headgear (Lewis et al., 2008; Watterson et al., 2005) and intra-subject variability. This personal variability could be due to the variation in performance and physiological factors such as small changes in nasal patency (de Boer & Bressmann, 2014; Lewis et al., 2008; van Doorn & Purcell, 1998). These natural variations of nasal patency over time may possibly explain the larger differences of the NSI 2.0 and its parameters in the long-term condition.
Table 6.1 Statistical measures of the test-retest reliability of the NSI 2.0 and its parameters in the short- and long-term condition for both adults and children: mean differences (mean_diff), standard deviations of the differences (SD_diff), intraclass correlation coefficient (ICC), standard error of measurement (SEM) and the minimal detectable difference to determine a confidence interval of 95% (MDD_95%)

<table>
<thead>
<tr>
<th>Reliability parameters</th>
<th>Short-term reliability</th>
<th>Long-term reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adults (n=40)</td>
<td>Children (n=29)</td>
</tr>
<tr>
<td>NSI Vowel /u/</td>
<td>-0.10</td>
<td>-0.07</td>
</tr>
<tr>
<td>Oral text</td>
<td>0.00%</td>
<td>-0.55%</td>
</tr>
<tr>
<td>VLHR /i/ 4.47*F0</td>
<td>-0.02dB</td>
<td>-0.92dB</td>
</tr>
<tr>
<td>Mean_diff</td>
<td>-0.10</td>
<td>-0.07</td>
</tr>
<tr>
<td>SD_diff</td>
<td>0.77</td>
<td>0.84</td>
</tr>
<tr>
<td>Oral text</td>
<td>2.16%</td>
<td>4.79%</td>
</tr>
<tr>
<td>VLHR /i/ 4.47*F0</td>
<td>2.72</td>
<td>0.15</td>
</tr>
<tr>
<td>ICC</td>
<td>0.92</td>
<td>0.89</td>
</tr>
<tr>
<td>Oral text</td>
<td>0.93%</td>
<td>0.83%</td>
</tr>
<tr>
<td>VLHR /i/ 4.47*F0</td>
<td>0.85</td>
<td>0.92%</td>
</tr>
<tr>
<td>SEM</td>
<td>0.55</td>
<td>0.59</td>
</tr>
<tr>
<td>Oral text</td>
<td>1.53%</td>
<td>3.39%</td>
</tr>
<tr>
<td>VLHR /i/ 4.47*F0</td>
<td>1.93dB</td>
<td>1.13%</td>
</tr>
<tr>
<td>MDD_95%</td>
<td>1.52</td>
<td>1.64</td>
</tr>
<tr>
<td>Oral text</td>
<td>4.23%</td>
<td>9.39%</td>
</tr>
<tr>
<td>VLHR /i/ 4.47*F0</td>
<td>2.97%</td>
<td>2.95%</td>
</tr>
<tr>
<td></td>
<td>5.34dB</td>
<td>7.20dB</td>
</tr>
</tbody>
</table>

†Because of heteroscedasticity, a logarithmic transformation was applied on these data, resulting in the absence of units for this parameter (Atkinson & Nevill, 1998). An antilog transformation was applied on the SEM (Y=e^x).
Short-term and long-term test-retest reliability of the NSI 2.0

Table 6.2 Cumulative frequencies (raw/%) of absolute differences in nasalance scores of the vowel /u/ and an oral text in adults and children for both, the short- and long-term condition.

<table>
<thead>
<tr>
<th>Nasalance difference score (%)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vowel /u/</td>
<td>Oral text</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adults (n=40)</td>
<td>Children (n=29)</td>
<td>Adults (n=19)</td>
<td>Children (n=11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short-term</td>
<td>Long-term</td>
<td>Short-term</td>
<td>Long-term</td>
<td>Short-term</td>
<td>Long-term</td>
<td>Short-term</td>
<td>Long-term</td>
</tr>
<tr>
<td>0</td>
<td>10/25</td>
<td>0/0</td>
<td>3/10</td>
<td>2/18</td>
<td>8/20</td>
<td>3/16</td>
<td>13/45</td>
<td>1/9</td>
</tr>
<tr>
<td>≤2</td>
<td>32/80</td>
<td>9/45</td>
<td>14/48</td>
<td>7/64</td>
<td>35/88</td>
<td>15/79</td>
<td>26/90</td>
<td>7/64</td>
</tr>
<tr>
<td>≤3</td>
<td><strong>36/90</strong></td>
<td>11/55</td>
<td>17/59</td>
<td>7/64</td>
<td><strong>39/98</strong></td>
<td><strong>17/90</strong></td>
<td>27/93</td>
<td>8/73</td>
</tr>
<tr>
<td>≤4</td>
<td>39/98</td>
<td>14/70</td>
<td>22/76</td>
<td>8/73</td>
<td>40/100</td>
<td>19/100</td>
<td>27/93</td>
<td><strong>11/100</strong></td>
</tr>
<tr>
<td>≤5</td>
<td>39/98</td>
<td>15/75</td>
<td>24/83</td>
<td>8/73</td>
<td>28/97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤6</td>
<td>39/98</td>
<td>17/85</td>
<td>24/83</td>
<td>8/73</td>
<td>28/97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤7</td>
<td>40/100</td>
<td>17/85</td>
<td><strong>26/90</strong></td>
<td>9/82</td>
<td>28/97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤8</td>
<td><strong>19/95</strong></td>
<td>26/90</td>
<td><strong>10/91</strong></td>
<td></td>
<td>28/97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤9</td>
<td>19/95</td>
<td>27/93</td>
<td>11/100</td>
<td></td>
<td>28/97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤10</td>
<td>19/95</td>
<td>27/93</td>
<td></td>
<td></td>
<td>29/100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤11</td>
<td>20/100</td>
<td>28/97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤12</td>
<td></td>
<td>28/97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤13</td>
<td></td>
<td>29/100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Data in bold represent the cutoff at which 90% of the nasalance score differences were included.

Application of Bland-Altman plots revealed that an adult with a higher, albeit normal nasalance score on the vowel /u/ showed higher long-term test-retest variability. This effect was not detected for the nasalance score of an oral text, VLHR, nor for the NSI 2.0 scores in the long-term condition. This is in contrast with the results reported by de Boer and Bressmann (2014) who found that participants with higher mean nasalance values for an oral text showed higher long-term test-retest variability. Additionally, a child with a high, albeit normal, nasalance score on an oral text showed higher short-term test-retest variability for this parameter. Although this observation was not found in the long-term condition, these findings may explain the higher test-retest variability in patients with cleft palate reported by Watterson and Lewis (2006), which has to be taken into account when interpreting successive measurements in patients. Further research should therefore focus on the test-retest reliability of the NSI 2.0 in patients presenting with hypernasality.

The results of the ICCs (Table 6.1) indicate an excellent relative consistency of the NSI 2.0 and its parameters. The smallest, albeit good, agreement was found for the parameter ‘VLHR of /i/’ in the short-term condition for children and in the long-term condition for both adults and children. Although the influence of noise and loudness variations was expected to be limited by using a relative rather than an absolute index (Lee et al., 2003), this lower ICC may indicate the importance of controlling for these influencing variables. Therefore, recordings are preferably made in a quiet room using the same mouth-microphone distance (i.e. 10cm).
Additionally, the participant can be asked to count aloud from one to three before sustaining the vowel /i/ to evoke his/her habitual loudness and pitch more easily. To the best of our knowledge, only Park et al. (2014) reported ICCs for nasalance values. Thirteen adults and 13 children with normal resonance read an oral text twice without replacement of the headset resulting in an ICC of 0.87. In the current study, a similar ICC of 0.92 was found for the oral text in both adults and children in the short-term condition, despite the replacement of the headset.

Because the application of only ICCs in reliability studies is not sufficient due to their dependency on the variability of the measured data and the absence of a consensus about their interpretation (Atkinson & Nevill, 1998; Costa-Santos et al., 2011), SEM and MDD were additionally explored in this study. These statistical parameters confirm the above-mentioned observations that the long-term variability is somewhat larger than the short-term variability and that the test-retest variability in adults and children is comparable. Using MDD, a 95% CI can be derived to reflect the interval in which 95% of the observations of a specific person will be found. The construction of such an interval is necessary when two measurements of the same patient with hypernasality are compared to detect any progress in this area due to surgery or, indirectly due to speech therapy, one of the ultimate goals of the NSI 2.0. If the second measurement remains within the limits of the 95% CI constructed around the first measurement, any difference between both measurements is most likely a result of personal variation or a measurement error. If the second measurement exceeds this interval, the observed difference is probably due to a clinically real or genuine difference. To illustrate the clinical implementation of this MDD, two clinical follow-up cases are presented below.

The first case is an 8-year-old boy born with bilateral cleft lip and palate. Closure of the lip was performed at the age of 6 months, closure of the palate at the age of 1 year. Speech therapy started at the age of 4 years. His articulation was characterized by glottal stops during the production of the plosives /t,d,k,p,b/ and nasal fricatives as a substitute for the fricatives /s,z,f,v/. His resonance was judged to be severely hypernasal. Speech therapy was provided four times a week. The boy consulted the university department of speech-language pathology each year to monitor his speech. His NSI 2.0 score yielded \(-14.7 \pm [13.20 \pm (0.0824 \times \text{nasalance } /u/ 77\%) – (0.260 \times \text{nasalance oral text } 56\%) – (0.242 \times \text{VLHR } /i/ 4.47 \times \text{F0Hz } 25.05 \text{dB})].\) Addition and subtraction with the MDD for children obtained from the long-term condition (i.e. MDD=2.68) resulted in a 95% CI of \([-17.4 – -12.0]\). Speech therapy was continued with a frequency of four times a week for one year with specific attention to the correct production of pressure consonants and the direction of the oral airflow during the production of fricatives. During the next follow-up assessment his NSI score yielded \(-7.9 \pm [13.20 – (0.0824 \times \text{nasalance } /u/ 62\%) – (0.260 \times \text{nasalance oral text } 43\%) – (0.242 \times \text{VLHR } /i/ 4.47 \times \text{F0Hz } 20.18 \text{dB})].\) As the score obtained during the second assessment exceeded the 95% CI derived from the first measurement score, this difference could be considered as a real difference. This difference may be due to the change of airflow as a more correct
articulation place of the fricatives was applied, so that hypernasality was indirectly improved by speech therapy. This improvement was also confirmed by the perceptual assessment in which the speech was judged as severely hypernasal during the first assessment and as moderately hypernasal during the second assessment.

The second case is a 7-year-old girl who consulted our department because of hypernasal speech due to a submucosal cleft palate. Before surgical closure of the cleft, her NSI score yielded -7.8 [13.20 – (0.0824 x nasalance /u/ 32%) – (0.260 x nasalance oral text 48%) – (0.242 x VLHR /i/ 4.47*F0Hz 24.43dB)]. Addition and subtraction of MDD resulted in a 95% CI of [-10.5 – -5.1]. Perceptually, her speech was judged as mildly to moderately hypernasal in combination with frequently observed nasal turbulence. Hypernasality was present at both high and low vowels. The cleft was surgically closed using the Sommerlad technique (Sommerlad, 2003). One month after surgery, a second assessment resulted in an NSI score of -5.9 [13.20 – (0.0824 x nasalance /u/ 61%) – (0.260 x nasalance oral text 33%) – (0.242 x VLHR /i/ 4.47*F0Hz 22.95dB)]. Although an improvement of the NSI score was observed, the score was included in the 95% CI derived from the NSI at the first assessment and therefore could not be interpreted as a real improvement of the nasal resonance, despite the decreased nasalance score on the oral text. Perceptually, her speech was judged after surgery as mildly to moderately hypernasal with hypernasality observed at high and low vowels. However, the nasal turbulence occurred only occasionally now, which may have explained the decreased nasalance score for the oral text. The equal NSI scores, on the other hand, represent the constant perceptual judgment of hypernasality and emphasize the clinical utility of the NSI 2.0 as a multiparametric index.

Considering the parameters of the NSI 2.0, reliability studies for the nasalance scores of an oral text or oral sentences are the only reported in literature. Reliability of the nasalance score of an oral text within one session with replacement of the headgear was found to vary up to 7% in 97% (Watterson et al., 2005) or 98% (Lewis et al., 2008) of the observations in adults without resonance disorders. For the long-term reliability, nasalance score differences between 5 and 9% were found in 95% of the observations in adults (de Boer & Bressman, 2014; Lewis et al., 2008; Whitehill, 2001) and differences between 6 to 9% were found in 94% of the observations in children (Lewis & Watterson, 2003; van Doorn & Purcell, 1998). Considering the short- and long-term condition in the present study, slightly smaller differences were found between two observations for both the adults and children. Regarding the adults, 100% of the observed differences were within 4% in both conditions (Table 6.2). Additionally, 97% of the observed differences in the children’s group were within 5% in the short-term condition and 100% of the differences were within 4% in the long-term condition (Table 6.2). A possible explanation for this smaller variability may be the language of the participants. Most previous reliability studies included English-speaking participants, whereas the subjects in the current study were all native speakers of Dutch. Van Lierde et al. (2001) reported significantly lower nasalance scores for an oral text in Dutch compared to Canadian English (Kavanagh et al., 1994), North-American English (Seaver et al., 1991) and
Spanish (Anderson, 1996). As lower nasalance scores may vary less compared to higher nasalance scores (de Boer & Bressmann, 2014), this may explain the smaller variability of the nasalance scores for an oral text found in the current study.

Nevertheless, only participants without resonance disorders were included in this study. As Watterson and Lewis (2006) stated, normal-speaking adults can be tested more reliably than patients. Higher test-retest variability was found in their study including patients with cleft palate (mean age 10y8m, 3y3m-26y), in which nasalance score differences up to 10% were reported for 94% of the consecutive measurements of an oral text in a short-term condition with replacement of the headgear. Therefore, as mentioned above, further research should focus on the reliability of the NSI 2.0 in patients with hypernasality.

Although the NSI 2.0 is proven to be a reliable instrument, it remains an indirect measure of only one particular aspect of speech, namely hypernasality. Consequently, no information can be provided about the intelligibility or acceptability of a patient’s speech based on the index score alone. As only oral stimuli are included in the index, additionally, no information can be derived about the presence of hyponasality or mixed nasality, which can be observed, for example, due to a septum deviation, narrow vestibule, maxillary retrusion or after speech improving surgery such as a pharyngeal flap (Fukushiro & Trindade, 2011; Kummer, 2011; Nellis, et al., 1992; Shprintzen, et al., 1979). These limitations emphasize the persistent need for reliable perceptual judgments to complement the interpretation of instrumental measurements results.

Conclusion

Based on the results of the current study, long-term test-retest variability of the NSI 2.0 and its parameters is slightly higher compared to short-term test-retest variability. This may be explained by the variation in personal performance and physiological changes of the nasal patency (de Boer & Bressmann, 2014; Lewis et al., 2008; van Doorn & Purcell, 1998), which may be larger when differences of a larger timespan are considered. The interval [NSI 2.0 ± MDD], i.e. [NSI 2.0±2.82] for adults and [NSI 2.0±2.68] for children, defines the 95% CI. Once a new obtained NSI 2.0 value lies outside this interval for a specific patient, the observed change is considered as due to genuine physiological changes. With an ICC of 0.84 in adults and 0.77 in children, the NSI 2.0 additionally shows an excellent relative consistency even in a long-term test-retest condition. Additionally, this study revealed that children can be tested as reliable as adults, since no significant differences in test-retest score differences are withheld. As the results of the current study are based on the assessment of normal-speaking participants, further research should investigate the reliability of the NSI 2.0 and its parameters in patients with perceived hypernasality. With the validation of this new index being the subject of ongoing research at this moment, the clinical implementation of this new, multiparametric approach will be completed soon and will consist a new, more powerful approach in the assessment of and treatment planning for individuals presenting with hypernasality.
Short-term and long-term test-retest reliability of the NSI 2.0

References


Short-term and long-term test-retest reliability of the NSI 2.0


CHAPTER 7

THE RELATIONSHIP BETWEEN THE NASALITY SEVERITY INDEX 2.0 AND PERCEPTUAL JUDGMENTS OF HYPERNASALITY


Abstract

Objective. The Nasality Severity Index 2.0 (NSI 2.0) forms a new, multiparametric approach in the identification of hypernasality. The present study aimed to investigate the correlation between the NSI 2.0 scores and the perceptual assessment of hypernasality.

Method. Speech samples of 35 patients, representing a range of nasality from normal to severely hypernasal, were rated by four expert speech-language pathologists using visual analogue scaling (VAS) judging the degree of hypernasality, audible nasal airflow (ANA) and speech intelligibility. Inter- and intra-listener reliability was verified using intraclass correlation coefficients. Correlations between NSI 2.0 scores and its parameters (i.e. nasalance score of an oral text and vowel /u/, voice low tone to high tone ratio of the vowel /i/) and the degree of hypernasality were determined using Pearson correlation coefficients. Multiple linear regression analysis was used to investigate the possible influence of ANA and speech intelligibility on the NSI 2.0 scores.

Results. Overall good to excellent inter- and intra-listener reliability was found for the perceptual ratings. A moderate, but significant negative correlation between NSI 2.0 scores and perceived hypernasality (r=-0.64) was found, in which a more negative NSI 2.0 score indicates the presence of more severe hypernasality. No significant influence of ANA or intelligibility on the NSI 2.0 was withheld.

Conclusion. As the NSI 2.0 correlates significantly with perceived hypernasality, it provides an easy-to-interpret severity score of hypernasality which will facilitate the evaluation of therapy outcomes, communication to the patient and other clinicians, and decisions for treatment planning, based on a multiparametric approach. However, research is still necessary to further explore the instrumental correlates of perceived hypernasality.

Key words. Nasality Severity Index 2.0; hypernasality; perceptual judgment; visual analogue scale
Learning outcomes. The reader will be able to (1) describe and discuss current issues and influencing variables regarding perceptual ratings of hypernasality; (2) describe and discuss the relationship between the Nasality Severity Index 2.0, a new multiparametric approach to hypernasality, and perceptual judgments of hypernasality based on visual analogue scale ratings; (3) compare these results with the correlations based on a single parameter approach and (4) describe and discuss the possible influence of audible nasal airflow and speech intelligibility on the NSI 2.0 scores.

Introduction

Speech perception is fundamentally perceptual in nature. A speech disorder only exists when it is recognized by the patient and/or others in the patient’s environment. Perceptual assessments of speech disorders by clinicians are based on this principle and remain the gold standard during diagnosis and evaluation of speech therapy and/or surgical intervention as it cannot be replaced yet by any instrumental assessment. Moreover, several authors consider perceptual assessments as the standard against which instrumental measurements must be validated (Kent, 1996; Keuning et al., 2004; Moll, 1964; Vogel et al., 2009). However, several variables can influence listeners’ perception of speech which may limit the reliability and validity of perceptual judgments. Kreiman et al. (1993) provide an overview of these factors. More specifically, individual differences due to experience, persons’ perceptual habits, biases, etc. which determine the listener’s internal standard can influence perceptual judgments. Additionally, task factors such as definitions of rating scales, listeners’ familiarity with the used scale and perceptual context (i.e. the ‘listener drift’ in which listeners rate speech as more severely disturbed when a moderately impaired speech sample follows a series of mildly impaired samples) may have a potential influence. Lastly, interaction between listeners’ and task factors, such as differences in the interpretation of the rating points of the used scale, can also influence listeners’ decisions. Furthermore, the discussion continues about the type of rating scale that has to be applied (Baylis et al., 2015; Brancamp et al., 2010; Schiavetti et al., 1983; Whitehill et al., 2002; Wuyts et al., 1999; Yiu & Ng, 2004; Zraick & Liss, 2000). Equal appearing interval (EAI) scales with clear description of the different grades are recommended in different perceptual assessment protocols for resonance disorders in patients with cleft palate (Henningsson et al., 2008; John et al., 2006; Sell, 2005; Sell et al., 1994, 1999; Sweeney & Sell, 2008). Moreover, EAI scaling was applied in 74% of the studies that included a perceptual assessment of cleft palate speech, as reported by Lohmander and Olsson (2004). However, recent studies suggested that resonance can be rated more reliably and validly by using ratio scales such as direct magnitude estimation (DME) or visual analogue scaling (VAS) (Baylis et al., 2015; Baylis et al., 2011; Whitehill et al., 2002; Zraick & Liss, 2000). Brancamp et al. (2010), on the other hand, reported no statistically significant differences between ratings of nasality based on EAI and DME scales. Nevertheless, listeners need adequate information about the terminology to describe nasality reliably (Kent et al., 1999; Whitehill, 2002). Additionally, inclusion of
reference samples and training sessions may improve consistency as the instable internal standard of the listener is then replaced by perceptual references (Kreiman et al., 1993).

To supplement perceptual evaluations of resonance disorders, several instrumental techniques are available (Bettens et al., 2014). A recent instrumental approach to identify hypernasality is the Nasality Severity Index version 2.0 (NSI 2.0) (Bettens et al., 2016), based on the originally developed NSI by Van Lierde et al. (2007). The NSI 2.0 includes different instrumental measurement techniques based on the optimal statistical discrimination of 35 children with perceived hypernasality and a control group of 50 children without resonance disorders, with sensitivity and specificity as the serving criteria. The index consists of a combination of three acoustic parameters: the nasalance value of the vowel /u/ and an oral text passage obtained by a Nasometer and the voice low tone to high tone ratio (VLHR) of a sustained vowel /i/ with a cutoff frequency of 4.47*F0Hz (originally described by Lee et al. (2003)). The formula yields NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%)) – (0.260 x nasalance oral text (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)). A cutoff score of zero was determined to discriminate patients from children without resonance disorders, resulting in a sensitivity of 92% (i.e. % correct identification of patients) and a specificity of 100% (i.e. % correct identification of controls), in which a score below zero is considered pathological. Additionally, the validity of the NSI 2.0 was proven to be high by applying the parameter results of an independent patient and control group on the derived formula (sensitivity 88%, specificity 89%, cutoff zero). Although the NSI 2.0 can discriminate patients with hypernasality from control children with a high sensitivity and specificity, the correlation between perceptual ratings of hypernasality and the index was not yet investigated. Additionally, perceptual evaluation was only performed by two investigators in the study by Bettens et al. (2016), which may be insufficient to rate the degree of hypernasality reliably.

Two of the three parameters of the NSI 2.0 are obtained by a Nasometer. This device, originally developed by Fletcher and Bishop (1973) and manufactured by KayPentax (NJ, Lincoln Park), measures the amount of nasalance by capturing the oral and nasal signal using two microphones on a sound separation plate which is placed under the nose of the patient. After bandpass filtering, the nasal signal is divided by the total signal of oral and nasal energy and multiplied by 100 to receive the nasalance score in percentage. This value represents an indirect measure of nasality. However, contradictory results have been reported about the correlation between the perceived degree of nasality and nasalance values. Correlation coefficients vary between 0.29 to 0.76 based on nasalance scores of oral stimuli determined by a Nasometer, with most authors reporting moderate correlations (Brancamp et al., 2010 – r=0.59 and r=0.63; Brunnegard et al., 2012 – r=0.49-0.76; Dalston et al., 1993 – r=0.73; Keuning et al., 2004 – r=0.54; Lewis et al., 2003 – r=0.29-0.57; Sweeney & Sell, 2008 – r=0.74; Watterson et al., 1993 – r=0.49). Differences in reported correlations may result from methodological differences between studies. Brunnegard et al. (2012), for example, reported different correlations depending on speech stimulus and experience of the listeners, in which the highest correlations are reported for the perceptual judgment of
Chapter 7

hypernasality based on spontaneous speech by speech-language pathologists and nasalance scores based on an oral stimulus. Another source of variance between studies is the inclusion of different patient populations. For example, the inclusion of patients who underwent pharyngeal flap surgery may decrease the correlation between nasalance and nasality (Hardin et al., 1992; Nellis et al., 1992). Furthermore, the presence of nasal airflow problems such as audible nasal emission and turbulence may influence nasalance scores (Karnell, 1995; Sweeney & Sell, 2008; Watterson et al., 1998). As a Nasometer cannot discriminate between acoustic energy from nasal resonance and energy from aerodynamic phenomena (such as audible nasal emission and turbulence), nasalance scores may be increased in patients with audible nasal airflow problems which can cause inconsistency with listeners’ judgments (Watterson et al., 1998). Inclusion of only consonants requiring low intra-oral pressure can prevent this ‘virtual’ increase because audible nasal airflow normally does not exist on those consonants. Nevertheless, good correlations have been reported between nasalance stimuli including high pressure consonants and perceptual judgments, even when patients with audible nasal airflow problems are included (Sweeney & Sell, 2008; Watterson et al., 1998). Moreover, Sweeney and Sell (2008) stated that a speech sample including all consonant types may have greater face validity because it represents spontaneous speech more adequately.

Considering VLHR, the third parameter included in the NSI 2.0, moderate to high correlations with perceptual judgments of hypernasality have been reported (Lee et al., 2009: $r_s=0.54-0.62$; Tsai et al., 2012: $r_s=0.79-0.81$). Although Vogel et al. (2009) did not find significant differences between various grades of hypernasality based on VLHR of /i/, Tsai et al. (2012) assigned these findings to an incorrect application of the analysis. Nevertheless, the analyses are mostly based on vowels only, which may not totally reflect hypernasality in connected speech.

With a high sensitivity and specificity, the NSI 2.0 accurately identifies hypernasality in patients with a history of cleft palate (Bettens et al., 2016). However, the final purpose of the NSI 2.0 is to provide an easy-to-interpret severity score of hypernasality to facilitate the evaluation of therapy outcomes, communication to the patient and other clinicians, and decisions for treatment planning. Therefore, the aim of the present study was to investigate the correlation between NSI 2.0 scores of children with hypernasality and perceptual assessment of hypernasality based on spontaneous speech and sentence repetition. A negative correlation between the NSI 2.0 and perceived severity of hypernasality is hypothesized, in which a more negative NSI 2.0 score correlates with the perception of more severe hypernasality. Moreover, due to its multidimensional approach, higher correlations with auditory perceived hypernasality ratings are hypothesized in comparison with a single parameter approach. Additionally, the possible influence of audible nasal airflow and speech intelligibility on the NSI 2.0 scores will be investigated, in which no influence of these variables is expected.
Method

This research was approved by the institutional review and ethical board of the Ghent University Hospital (EC/2012/049).

Listeners

Four speech-language pathologists with at least 5 years of experience in rating hypernasality served as listeners in this study. The different aspects related to these listeners are summarized in Table 7.1.

Speech samples

Speech samples were collected from 35 children between 4 and 15 years old (mean age 7.3y, SD 2.67), representing a range of nasality from normal to severely hypernasal. All patients consulted the department of speech and language pathology at the Ghent University Hospital in Belgium with a complaint of hypernasal speech due to a variety of pathologies or during a follow-up period after palatal closure (Table 7.2). The inclusion criteria to participate in this study were being a native speaker of Dutch, living in Flanders (the northern part of Belgium), and being able to produce the required speech sample with good cooperation. Patients suffering from a cold or allergy outburst at the moment of testing or presenting with hyponasal resonance, a pharyngeal flap, learning disabilities greater than mild, dysarthria or dyspraxia were excluded from the study. All three patients included in the category ‘velopharyngeal mislearning’ showed hypernasality on vowels, whether or not in combination with nasal fricatives during the production of specific consonants. Videofluoroscopic and/or nasoendoscopic assessment showed the possibility to properly close the velopharyngeal mechanism, however, velopharyngeal closure was not continuously observed, resulting in the presence of mild hypernasality.

Two separate speech samples, one including spontaneous speech and one including sentence repetition, were collected for each child. Sixty-five syllables (i.e. the length of the smallest available sample) were selected to compose the spontaneous speech sample for each child, resulting in speech samples with a similar length in terms of number of syllables. The second speech sample included the repetition of the Dutch translation of 12 oral and 3 nasal sentences derived from the MacKay-Kummer Simplified Nasometric Assessment Procedures (SNAP) test (Kummer, 2005; MacKay & Kummer, 1994). All samples were video-recorded using a Sony HDR-CX280 camera with integrated high-quality microphone in a quiet room at the clinical department of the Ghent University Hospital. To limit listener bias, all samples were converted to audio samples using audio converter software (Freemake audio converter, version 1.1.0.66) at a sampling frequency of 48kHz.
Table 7.1 Listener characteristics.

<table>
<thead>
<tr>
<th>Listener</th>
<th>Gender</th>
<th>Mother tongue</th>
<th>PTA (dB HL)</th>
<th>Years of experience</th>
<th>Working setting</th>
<th>Amount of patients with resonance disorders seen each year</th>
<th>Experience with EAI</th>
<th>Experience with VAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>Dutch</td>
<td>5</td>
<td>&gt;10y</td>
<td>University hospital, craniofacial team</td>
<td>51-100</td>
<td>Resonance, voice, speech intelligibility</td>
<td>Voice</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>Dutch</td>
<td>10</td>
<td>&gt;10y</td>
<td>University hospital rehabilitation center</td>
<td>31-50</td>
<td>Resonance, voice, speech intelligibility</td>
<td>Voice, speech intelligibility</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>Dutch</td>
<td>3</td>
<td>5y</td>
<td>University hospital, craniofacial team</td>
<td>51-100</td>
<td>Resonance, voice, speech intelligibility</td>
<td>Not applicable</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>Dutch</td>
<td>-2</td>
<td>&gt;10y</td>
<td>General hospital, craniofacial team</td>
<td>11-30</td>
<td>Resonance, voice, speech intelligibility</td>
<td>Voice</td>
</tr>
</tbody>
</table>

PTA = pure threshold audiometry based on the arithmetic mean of the threshold on 500Hz, 1kHz and 2kHz; EAI = equal appearing interval scale; VAS = visual analogue scale
Table 7.2 Diagnosis of the included patients (based on Trost-Cardamone (1989)).

<table>
<thead>
<tr>
<th>Classification</th>
<th>Number of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolated cleft palate</td>
<td>15 43</td>
</tr>
<tr>
<td>Bilateral cleft lip and palate</td>
<td>7 20</td>
</tr>
<tr>
<td>Unilateral cleft lip and palate</td>
<td>4 11</td>
</tr>
<tr>
<td>Submucosal cleft palate</td>
<td>1 3</td>
</tr>
<tr>
<td>Velopharyngeal insufficiency after adenotomy</td>
<td>2 6</td>
</tr>
<tr>
<td>Velopharyngeal insufficiency after adenotonsillectomy</td>
<td>1 3</td>
</tr>
<tr>
<td>Velopharyngeal insufficiency due to velo-cardio-facial syndrome</td>
<td>2 6</td>
</tr>
<tr>
<td>Velopharyngeal mislearning</td>
<td>3 8</td>
</tr>
</tbody>
</table>

**Instrumental assessment**

After collecting the speech sample, all parameters of the NSI 2.0 were determined following the protocol described by Bettens et al. (2016). Nasalance scores of the vowel /u/ and an oral text were collected using a KayPentax Nasometer II model 6450. After calibration in a quiet room and according to the manufacturer’s instruction manual, each child was asked to repeat the vowel /u/ four times while the headset of the Nasometer was placed against the upper lip beneath the nostrils. Additionally, the child was asked to read or repeat an oral text passage devoid of nasal consonants, originally developed by Van de Weijer and Slis (1991), at a comfortable loudness and pitch. To determine VLHR of /i/, the child was asked to sustain the vowel /i/ for at least 2 seconds in front of a unidirectional microphone (Samson, C01U) placed at a distance of approximately 10cm from the child’s mouth. Recordings were made using PRAAT software, version 5.3.78 (Boersma & Weenink, 2014), at a sampling frequency of 44100Hz. Analyses were performed on a 0.5s fragment, extracted using a Hamming window and starting from at least 0.5s after the voice onset to exclude any possible influence of this onset. Based on a cutoff frequency of 4.47*F0Hz, the spectrum was divided into a low frequency (65Hz to cutoff) and high frequency (cutoff to 8000Hz) spectral region according to Lee et al. (2003). After summation of the power of each frequency region, VLHR was determined as the power ratio of the low frequency to the high frequency spectrum in accordance with Lee et al. (2003; 2006; 2009). The analysis was completed using a PRAAT script to facilitate the proceeding (see appendix).

After collecting all parameters of the NSI 2.0, they were entered into the formula NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%)) – (0.260 x nasalance oral text (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)) to compute the NSI 2.0 score of each child.

**Perceptual ratings**

All listeners were invited for a listening session at the department of speech, language and hearing sciences of the Ghent University. At the beginning of the rating session, information was given by the first author about rating procedure, rating scale and terminology of the categories that had to be rated. A short training session was given using
different examples of normal resonance, hypernasality and hypernasality with audible nasal airflow to improve consistency. Each listener was asked to rate the degree of hypernasality, audible nasal airflow, and speech intelligibility of the examples after which the ratings were discussed. Hypernasality was defined as “any abnormal increase in nasal resonance during speech production which is most easily perceived on vowels and voiced consonants” and audible nasal airflow was defined as “any abnormal or inappropriate audible escape of air from the nasal cavity accompanying the production of oral pressure consonants” according to the definitions provided by John et al. (2006). Speech intelligibility was defined as “the degree to which the speaker’s message can be understood by the listener” according to the definition provided by Henningsson et al. (2008). This latter aspect was only rated on the spontaneous speech samples.

During the training session, ratings were performed using VAS by placing a mark on a 100mm bar. Per sample, three bars were provided including the label ‘absent’ or ‘normal’ at the left end and the label ‘severely distorted’ or ‘frequently noted’ at the right end (Baylis et al., 2015). The example samples were not included in the study samples. After completion of the training session, each listener received a standard pair of over-ear headphones and the blinded audio samples in a randomized sequence to exclude any order effect. To verify inter-rater reliability, 26% of the samples (9/35) were repeated. Careful control of the randomization took place by the first author so that two identical samples did not succeed immediately. Each sample could be listened to as often as needed, however, once the listener moved on to the next sample, the listener was asked not to return to a previous one. The same rating procedure as during the training session was followed. To exclude any order effect, two raters were at random assigned to rate the samples based on spontaneous speech, while the other two listeners rated the samples based on sentences first. A break was inserted after completing all speech samples in the same condition (sentences or spontaneous speech), after which the speech samples based on the opposite condition were rated. Any questions were answered by the first author during the rating procedure.

Data analysis

Inter-listener reliability was determined based on the intraclass correlation coefficient (ICC) using a two-way fixed model with consistency agreement (ICC (3,k) following the classification of Shrout and Fleiss (1979)) using SPSS software version 22.0 (IBM Corp., Armonk, NY). Intra-listener reliability was verified using a two-way fixed model with consistency agreement (ICC (3,1)). To verify the correlation between degree of perceptually observed hypernasality based on mean VAS scores of the four listeners and the NSI 2.0 scores and its parameters, Pearson correlation coefficients were determined. Before proceeding, assumptions of normality and linearity were verified and accepted for all variables, including mean VAS scores of hypernasality based on the sentences samples and spontaneous speech samples, NSI 2.0 and its parameters. Furthermore, multiple regression analyses were used to evaluate possible additional influence of audible nasal airflow and
speech intelligibility on the NSI 2.0 scores. Therefore, a hierarchical regression was applied in which the predictor ‘mean VAS score of hypernasality’ was forcibly entered into the model after which the predictors ‘audible nasal airflow’ and ‘speech intelligibility’ were entered using a backward stepwise approach. This procedure was performed for both the scores obtained from the spontaneous speech samples (predictors entered: hypernasality, audible nasal airflow and speech intelligibility) and those based on the sentences samples (predictors entered: hypernasality and audible nasal airflow). For each model, the correlation between the predictor variables was estimated by collinearity statistics and expressed by the variance inflation factor (VIF). VIF values below 5 were considered to imply no inter-correlations between predictor variables (Menard, 2002).

As no control subjects were included in the present study, no cutoff scores for instrumental measurement data to discriminate children with perceived hypernasality from children with normal resonance could be derived based on the data of this study. Therefore, cutoff scores for the NSI 2.0 and its parameters were determined based on the data obtained from the study by Bettens et al. (2016), who included 35 children with perceived hypernasality (mean age 7.6y, SD 3.10) and 50 control children with normal resonance (mean age 8.5y, SD 2.13). Cutoff scores were determined using receiver operating characteristic (ROC) curves with sensitivity and specificity as the serving criteria resulting in the following cutoff scores: 0 for the NSI 2.0 (sensitivity: 92%, specificity: 100%), 20% for the nasalance score of an oral text (sensitivity: 83%, specificity: 98%), 25% for the nasalance score of /u/ (sensitivity: 89%, specificity: 96%) and 24.83dB for VLHR of /i/ (sensitivity: 78%, specificity: 80%). These cutoff scores were applied to determine the amount of correctly identified children with perceived hypernasality (i.e. a mean VAS score above 10 mm) based on the instrumental measurement results in the present study.

**Results**

**Listener reliability**

Table 7.3 shows the ICCs and 95% confidence intervals (95% CI) for the inter-rater reliability of the rated categories. Additionally, an interpretation of the levels of agreement is provided for the average-measures ICCs regarding the guidelines by Cichette (1994). Good agreement was found between the listeners for the category ‘occurrence of audible nasal airflow’. Moreover, excellent agreement was found for the categories ‘hypernasality’ and ‘speech intelligibility’. Single-measures ICCs and their 95% CIs for the intra-rater reliability are provided in Table 7.4. Overall, good to excellent levels of agreement were found for all parameters, however, only a fair agreement was found for the judgment of hypernasality based on the spontaneous speech samples by listener 4. Moreover, the ratings of listener 2 on speech intelligibility agreed just poorly.
Table 7.3 Inter-listener reliability for perceptual ratings of hypernasality, audible nasal airflow (ANA) and speech intelligibility based on spontaneous speech and sentences using a visual analogue scale.

<table>
<thead>
<tr>
<th></th>
<th>Single-Measures ICC</th>
<th>95% CI</th>
<th>Average-Measures ICC</th>
<th>95% CI</th>
<th>Level of agreement*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypernasality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous</td>
<td>0.63</td>
<td>0.47-0.77</td>
<td>0.87</td>
<td>0.78-0.93</td>
<td>Excellent</td>
</tr>
<tr>
<td>Sentences</td>
<td>0.69</td>
<td>0.55-0.81</td>
<td>0.90</td>
<td>0.83-0.94</td>
<td>Excellent</td>
</tr>
<tr>
<td><strong>ANA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous</td>
<td>0.41</td>
<td>0.24-0.60</td>
<td>0.74</td>
<td>0.56-0.86</td>
<td>Good</td>
</tr>
<tr>
<td>Sentences</td>
<td>0.35</td>
<td>0.18-0.54</td>
<td>0.68</td>
<td>0.47-0.82</td>
<td>Good</td>
</tr>
<tr>
<td><strong>Speech intelligibility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous</td>
<td>0.75</td>
<td>0.63-0.85</td>
<td>0.92</td>
<td>0.87-0.96</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

*based on Cicchetti (1994): excellent: 0.75-1.00, good: 0.60-0.74, fair: 0.40-0.59, poor: <0.40

**Correlation between hypernasality ratings and instrumental measurements scores**

Figures 7.1 and 7.2 provide scatter plots including the Pearson correlation coefficients between the instrumental measurement scores (i.e. NSI 2.0 scores and its parameters) and the mean hypernasality ratings using VAS based on spontaneous speech samples and sentences. A general trend can be observed in which lower NSI 2.0 scores and higher nasalance scores for an oral text and the vowel /u/ correspond with higher perceptual ratings of hypernasality. Consequently, a moderate (Hinkle et al., 2003), but significant, negative correlation was found between the NSI 2.0 scores and the hypernasality ratings, in which a somewhat higher correlation could be observed for the ratings based on spontaneous speech samples (Pearson correlation coefficient: r=-0.64, p<0.001) vs. ratings based on sentences (Pearson correlation coefficient: r=-0.58, p<0.001). It is apparent from these figures that the correlation values of the nasalance scores of the oral text (spontaneous speech, r=0.63, p<0.001; sentences, r=0.57, p<0.001) are comparable, although opposite in direction, with those of the NSI 2.0 scores. Additionally, a low, although significant, correlation was observed for the nasalance scores of /u/ (spontaneous speech, r=0.48, p=0.004; sentences, r=0.37, p=0.031), while no correlation was found for VLHR scores of /i/ (spontaneous speech, r=-0.02, p=0.909; sentences, r=0.03, p=0.849).
### Table 7.4 | Intra-listener reliability for perceptual ratings of hypernasality, audible nasal airflow (ANA) and speech intelligibility based on spontaneous speech and sentences using a visual analogue scale.

<table>
<thead>
<tr>
<th>Speech sample</th>
<th>Listener 1</th>
<th>95% CI</th>
<th>Listener 2</th>
<th>95% CI</th>
<th>Listener 3</th>
<th>95% CI</th>
<th>Listener 4</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hypernasality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous</td>
<td>0.60</td>
<td>-0.05-0.89</td>
<td>0.90</td>
<td>0.63-0.98</td>
<td>0.93</td>
<td>0.73-0.98</td>
<td>0.42</td>
<td>-0.34-0.85</td>
</tr>
<tr>
<td>Sentences</td>
<td>0.78</td>
<td>0.28-0.95</td>
<td>0.87</td>
<td>0.54-0.97</td>
<td>0.95</td>
<td>0.78-0.99</td>
<td>0.92</td>
<td>0.67-0.98</td>
</tr>
<tr>
<td><strong>ANA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous</td>
<td>0.94</td>
<td>0.77-0.99</td>
<td>0.96</td>
<td>0.83-0.99</td>
<td>0.92</td>
<td>0.70-0.98</td>
<td>0.93</td>
<td>0.73-0.98</td>
</tr>
<tr>
<td>Sentences</td>
<td>0.97</td>
<td>0.86-0.99</td>
<td>0.69</td>
<td>0.10-0.92</td>
<td>0.78</td>
<td>0.30-0.95</td>
<td>0.88</td>
<td>0.56-0.97</td>
</tr>
<tr>
<td><strong>Speech intelligibility</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneous</td>
<td>0.98</td>
<td>0.91-1.00</td>
<td>0.37</td>
<td>-0.34-0.81</td>
<td>0.96</td>
<td>0.84-0.99</td>
<td>0.90</td>
<td>0.63-0.98</td>
</tr>
</tbody>
</table>
**Figure 7.1** Scatter plots of the NSI 2.0 and its parameters in function of the mean rating of hypernasality using a visual analogue scale (VAS) based on spontaneous speech samples (n=35). Cutoff scores are presented to visualize sensitivity and overall efficacy.

*Speech samples that were not correctly classified by the NSI 2.0 based on mean VAS scores for hypernasality applying cutoff scores of NSI 2.0<0 and VAS>10mm to represent hypernasality.

a. NSI 2.0 in function of hypernasality ratings (VAS), r=-0.64, p<0.001; sensitivity 97%, overall efficacy 91%, cutoff NSI 2.0<0, cutoff VAS>10mm

b. Nasalance score of an oral text in function of hypernasality ratings (VAS), r=0.63, p<0.001; sensitivity 90%, overall efficacy 89%, cutoff nasalance oral text>20%, cutoff VAS>10mm

c. Nasalance score of /u/ in function of hypernasality ratings (VAS), r=0.48, p=0.004; sensitivity 94%, overall efficacy 89%, cutoff nasalance /u/>25%, cutoff VAS>10mm

d. Voice low tone to high tone ratio of /i/ in function of hypernasality ratings (VAS), r=-0.02, p=0.909; sensitivity 61%, overall efficacy 60%, cutoff VLHR /i/>24.83dB, cutoff VAS>10mm
Relationship between NSI 2.0 and perceptual judgments of hypernasality

Figure 7.2 Scatter plots of the NSI 2.0 and its parameters in function of the mean rating of hypernasality using a visual analogue scale (VAS) based on sentences samples (n=35). Cutoff scores are presented to visualize sensitivity and overall efficacy.

†Speech samples that were not correctly classified by the NSI 2.0 based on mean VAS scores for hypernasality applying cutoff scores of NSI 2.0<0 and VAS>10mm to represent hypernasality.

a. NSI 2.0 in function of hypernasality ratings (VAS), r=-0.58, p<0.001; sensitivity 100%, overall efficacy 89%, cutoff NSI 2.0<0, cutoff VAS>10mm
b. Nasalance score of an oral text in function of hypernasality ratings (VAS), r=0.57, p<0.001; sensitivity 93%, overall efficacy 86%, cutoff nasalance oral text>20%, cutoff VAS>10mm
c. Nasalance score of /u/ in function of hypernasality ratings (VAS), r=0.37, p=0.031; sensitivity 96%, overall efficacy 86%, cutoff nasalance /u/>25%, cutoff VAS>10mm
d. Voice low tone to high tone ratio of /i/ in function of hypernasality ratings (VAS), r=0.03, p=0.849; sensitivity 64%, overall efficacy, 63%, cutoff VLHR /i/>24.83dB, cutoff VAS>10mm
Sensitivity and overall efficacy of the instrumental measurements based on ratings of spontaneous speech

As can be derived from figure 7.1a, 97% (30/31) of the spontaneous speech samples with perceived hypernasality were correctly classified by the NSI 2.0, i.e. 97% of the speech samples with a mean VAS score above 10mm had an NSI 2.0 score below zero (Bettens et al., 2016). In total 32 of the 35 samples were correctly classified by the NSI 2.0, resulting in an overall efficacy of 91%. Based on the nasalance scores of an oral text, using a cutoff score of 20%, 90% (28/31) of the hypernasal samples and 89% (31/35) of all samples in total were correctly classified (Figure 7.1b). Furthermore, based on the nasalance score of /u/, 94% (29/31) of the hypernasal speech samples and 89% (31/35) of all samples were correctly identified when applying a cutoff of 25% (Figure 7.1c). Based on VLHR of /i/, 61% (19/31) of the hypernasal speech samples and 60% (21/35) of all samples were correctly classified using a cutoff of 24.83dB (Figure 7.1d). As only four speech samples were judged with a negligible amount of hypernasality (mean VAS score ≤ 10mm), the overall efficacy of the instrumental measurements should be interpreted with care.

Sensitivity and overall efficacy of the instrumental measurements based on ratings of sentences

As can be derived from figure 7.2a, 100% (28/28) of the spontaneous speech samples with perceived hypernasality were correctly classified by the NSI 2.0. In total 31 of the 35 samples were correctly classified by the NSI 2.0, resulting in an overall efficacy of 89%. Based on the nasalance scores of an oral text, using a cutoff score of 20%, 93% (26/28) of the hypernasal samples and 86% (30/35) of all samples in total were correctly classified (Figure 7.2b). Furthermore, based on the nasalance score of /u/, 96% (27/28) of the hypernasal speech samples and 86% (30/35) of all samples were correctly identified when applying a cutoff of 25% (Figure 7.2c). Based on VLHR of /i/, 64% (18/28) of the hypernasal speech samples and 63% (22/35) of all samples were correctly classified using a cutoff of 24.83dB (Figure 7.2d). As only seven speech samples were judged without or with a negligible amount of hypernasality (mean VAS score ≤ 10mm), the overall efficacy of the instrumental measurements should be interpreted with care.

Influence of perceived audible nasal airflow and speech intelligibility on the NSI 2.0 scores, based on ratings of spontaneous speech

Audible nasal airflow was observed in 80% (28/35) of the patients (i.e. VAS score > 10mm) (mean 32mm, SD 23.5, range 0-82mm) based on the ratings of spontaneous speech. Additionally, deviating speech intelligibility was observed in 80% (28/35) of the patients (i.e. VAS score > 10mm) (mean 50mm, SD 32.0, range 0-97mm). Multiple linear regression analysis with a forced enter of the predictor ‘mean VAS score of hypernasality’ and a backward stepwise approach for the predictors ‘audible nasal airflow’ and ‘speech intelligibility’ revealed the selection of only the predictor ‘mean VAS score of hypernasality’
in the model to predict the NSI 2.0 scores \((t(33)=-4.776, p<0.001; B=-0.13, SE B=0.03, 95\% CI B=[-0.18;-0.07]; \beta=-0.64; R^2 \text{ adjusted}=0.39)\). As the predictors ‘audible nasal airflow’ and ‘speech intelligibility’ were not retained in the model, no significant contribution of these variables to the NSI 2.0 scores can be presumed.

**Influence of perceived audible nasal airflow on the NSI 2.0 scores, based on ratings of sentences**

Audible nasal airflow was observed in 83% (29/35) of the patients (i.e. VAS score > 10mm) (mean 36mm, SD 23.1, range 0-82mm) based on the ratings of sentences. Multiple linear regression analysis with a forced enter of the predictor ‘mean VAS score of hypernasality’ and a backward stepwise approach of the predictor ‘audible nasal airflow’ revealed the selection of only the predictor ‘mean VAS score of hypernasality’ in the model to predict the NSI 2.0 scores \((t(33)=-4.034, p<0.001; B=-0.10, SE B=0.02, 95\% CI B=[-0.15;-0.05]; \beta=-0.58; R^2 \text{ adjusted}=0.31)\). As the predictor ‘audible nasal airflow’ was not retained in the model, no significant contribution of this variable to the NSI 2.0 scores can be presumed.

**Discussion**

The purpose of the present study was to verify the correlation between the NSI 2.0, a new multiparametric instrumental approach to hypernasality, and the degree of perceived hypernasality. Additionally, the possible influence of audible nasal airflow and speech intelligibility on the NSI 2.0 was investigated. Based on the results of this study, the NSI 2.0 correlates moderately with the perceptual ratings of hypernasality \((r=0.64)\), in which a more negative NSI 2.0 score indicates the presence of more severe hypernasality. A comparable, but positive, correlation was found for nasalance scores of an oral text \((r=0.63)\), which is in agreement with correlations reported for this variable in literature (Brancamp et al., 2010 – \(r=0.59\) and \(r=0.63\); Brunnegard et al., 2012 – \(r=0.49-0.76\); Dalston et al., 1993 – \(r=0.73\); Keuning et al., 2004 – \(r=0.54\); Lewis et al., 2003 – \(r=0.29-0.57\); Sweeney & Sell, 2008 – \(r=0.74\); Watterson et al., 1993 – \(r=0.49\)). However, this finding is in contrast with the hypothesis that a multiparametric approach should correlate better with the degree of perceived hypernasality than a single parameter approach. A possible explanation may be that the additional parameters included in the NSI 2.0 (i.e. nasalance score of /u/ and VLHR of /i/) are only based on single vowels which may explain the low correlations with perception of the degree of hypernasality in continuing speech. Nevertheless, inclusion of these parameters into the NSI 2.0 contributes to a better identification of patients with hypernasality in comparison with a single parameter approach based on nasalance scores of an oral text alone, as reported by Bettens et al. (2016) (sensitivity NSI 2.0: 92%, sensitivity nasalance score oral text: 81%) and confirmed by the present study (sensitivity NSI 2.0: 97% and 100%, sensitivity nasalance score oral text: 90% and 93%). Moreover, a high overall efficacy was found in the present study for the NSI 2.0 (91%, 89%), indicating a good overall classification of speech samples with or without the presence of hypernasality. However, overall efficacy
scores have to be interpreted with care as only a small amount of samples was rated without or with a minimum of hypernasality.

Contrary to expectations based on previous research (Lee et al., 2009), no correlation was found between VLHR of /i/ and hypernasality ratings. Lee et al. (2009) reported correlations between 0.54 and 0.62 between VLHR and the perceptual evaluation of the single vowels /a, i, u, ë, ë, à/. However, no separate correlations for the vowel /i/ were described. As their perceptual judgments were only based on single vowels and most of the patients were judged with moderate-severe or severe hypernasality, this may explain the higher correlation found in their study. Additionally, Vogel et al. (2009) also reported about the inability of VLHR to discriminate between different grades of hypernasality, although Tsai et al. (2012) assigned this finding to an incorrect application of the procedure to calculate VLHR. Despite the lack of a correlation between VLHR of /i/ and degree of hypernasality, the weight of this parameter in the NSI 2.0 equation is comparable with the weight of the parameter ‘nasalance of an oral text’ (Bettens et al., 2016). A possible explanation for this contradiction may be the difference in the applied rating system and subsequent statistical analysis. To derive the NSI 2.0 equation, two listeners judged samples of spontaneous speech using the CAPS-A rating system (John et al., 2006), i.e. an EAI scale with grades from 0 to 4 to indicate the degree of hypernasality (Bettens et al., 2016). Samples with a consensus score of 0 were classified as ‘absence of hypernasality’, samples with a score of 1 or higher were classified as ‘presence of hypernasality’. Consequently, a binary logistic regression analysis was applied to derive the NSI 2.0 equation based on this classification. As VLHR of /i/ contributed to this classification, this parameter was selected by the logistic regression analysis (Bettens et al., 2016). As no correlation with degree of hypernasality was taken into account in that analysis, this may explain the contradictory findings regarding VLHR of /i/ in the current study. This hypothesis could be verified by comparing the influence of this parameter in a multiple regression analysis using the VAS score of degree of hypernasality as dependent variable and a binary logistic regression analysis based on the presence or absence of hypernasality. As an instrumental measurement requires both, a high correlation with the perceived degree of hypernasality and a high discriminatory value, VLHR of /i/ may still have a significant contribution to the NSI 2.0 equation.

As some authors attribute disagreements between perceptual assessment of hypernasality and instrumental measurements to the presence of audible nasal airflow (Karnell, 1995; Sweeney & Sell, 2008; Watterson et al., 1998) or misarticulations (Garcia et al., 2014), a multiple regression analysis was performed, including ‘hypernasality’, ‘audible nasal airflow’ and ‘speech intelligibility’ as predictors for the NSI 2.0 score. With only the degree of rated hypernasality withheld in the model, no significant influence of audible nasal airflow and speech intelligibility on the NSI 2.0 scores was retained, although the presence of audible nasal airflow and deviating speech intelligibility was observed in 80% of the patients. This is in accordance with Keuning et al. (2002) who found low correlations between nasalance
scores of an oral text and the amount of perceived audible nasal emission ($r=0.18$) and intelligibility ($r=0.34$).

Although 39% of the variance of the NSI 2.0 scores could be explained by a multiple regression model based on hypernasality ratings alone, the authors assume that the NSI 2.0 scores are possibly influenced by other variables. In the literature, acoustic restrictions of the Nasometer are suggested to explain disagreements with perceptual assessments of hypernasality (Keuning et al., 2002; Watterson et al., 1993). As nasalance is determined at a center frequency of 500Hz using a bandwidth of 300Hz (Fletcher & Bishop, 1973) while the acoustic effects of hypernasality are not restricted to this frequency range (Watterson et al., 1993), nasalance scores may represent only a part of perceived hypernasality (Keuning et al., 2002). To reduce this limitation, VLHR of the vowel /i/, which is based on a ratio that includes a wider frequency spectrum of speech, namely 65 to 8000Hz, was included in the analysis to compose the multiparametric NSI 2.0 (Bettens et al., 2016). However, no correlation between this single parameter and perceptual ratings of hypernasality was found in the present study. As mentioned above, a possible explanation may be that the VLHR score is based on a sustained vowel only, while perceptual assessment was based on spontaneous speech and sentence repetition. Determination of VLHR of /i/ extracted from words or sentences may result in higher correlations, but further research is needed to confirm this. Recently, de Boer and Bressmann (2015) proposed the application of long-term average spectra (LTAS) to determine the presence of oral-nasal balance disorders in connected speech. Such as VLHR, LTAS is based on the presence of extra- and antiresonance in the spectrum of hypernasal speech, including a broader frequency spectrum compared to the Nasometer. Inclusion of this new technique as a parameter in the NSI may result in higher correlations with perceptual judgments. However, although their first results based on simulations of hyper- and hyponasality are promising (de Boer & Bressmann, 2015), further research including patient based analyses is necessary to confirm the validity and applicability of this technique.

Another influencing variable may be the consistency of hypernasality (Sweeney & Sell, 2008), in which the degree of hypernasality may vary depending on the applied speech sample. In the present study, speech samples for perceptual assessment (i.e. spontaneous speech and sentence repetition) and NSI 2.0 calculations (i.e. oral text and sustained vowels) differed. However, Brunnegard et al. (2012) reported the highest correlation between perceptual ratings based on spontaneous speech and nasalance scores of oral sentences. This is in accordance with the results of the current study in which somewhat higher correlations were found between instrumental measurements and ratings based on spontaneous speech compared with those based on sentence repetition. Additionally, more samples based on sentence repetition were rated with a minimum of hypernasality, i.e. mean VAS score ≤ 10mm (n=7), in comparison with those including spontaneous speech (n=4). However, instrumental measurement data agreed less with this perception as can be concluded from the lower overall efficacy levels for the classification of speech samples based on sentences.
This may suggest, following Brunnegard et al. (2012), that listeners may rate hypernasality as good, if not better, in a more realistic speech sample than in standard stimuli. Furthermore, Sweeney and Sell (2008) also mentioned high correlations between the perceptual assessment based on a comprehensive speech sample and nasalance scores of high pressure sentences which are comparable with the oral text included in the NSI 2.0.

At last, task factors (e.g. familiarity with the used rating scale and perceptual context) and individual differences (e.g. experience, attention lapses, fatigue, mistakes) could have influenced the validity and reliability of the perceptual ratings (Kreiman et al., 1993). Despite the widely use of EAI scaling to assess cleft palate speech (Lohmander & Olsson, 2004), VAS was applied for the perceptual assessment of hypernasality, audible nasal airflow and speech intelligibility in the present study. VAS was opted for because of its ease of use, its ease of analysis and its wider range of statistical analysis options compared to EAI scale rating (Baylis et al., 2015). Moreover, recent studies (Baylis et al., 2015; Whitehill et al., 2007) confirmed the validity and reliability of the VAS resulting in a valid alternative method to EAI ratings for the assessment of cleft palate speech. Although the limited experience of the listeners in using this rating method, an overall good to excellent inter- and intra-listener reliability was found for the rated variables in the present study. Furthermore, speech samples were offered in a random order to exclude the influence of perceptual context. Additionally, the influence of individual differences was minimized by providing a training session based on reference samples. However, perceptual ratings were still based on personal judgments which remained sensitive to possible bias.

Although the NSI 2.0 correlates moderately with the perceived amount of hypernasality, it remains an indirect measure of only this particular aspect of speech. Although hypernasality affects speech intelligibility and acceptability more than hyponasality does and therefore is clinically more relevant (Shprintzen et al., 1979), some authors highlight the need to identify the amount of hyponasality and mixed nasality, which can be observed, for example, due to a septum deviation, narrow vestibule, maxillary retrusion or after speech improving surgery such as a pharyngeal flap (Fukushiro & Trindade, 2011; Kummer, 2011; Nellis et al., 1992; Shprintzen et al., 1979). Additionally, the patients included in the present study were carefully selected in order to isolate the influence of individual components within the experimental design. In more complex speech presentations, e.g. patients with mixed nasality, a pharyngeal flap, greater than mild global developmental delays or motor speech disorders, the interpretation of the NSI 2.0 may become more complicated. These limitations emphasize the persistent need to explore additional instrumental correlates of resonance disorders, as well as the need for reliable perceptual judgments to complement the interpretation of instrumental measurements results.

In conclusion, the NSI 2.0 correlates significantly with the perceptual ratings of hypernasality in which a score below 0 indicates the presence of hypernasality and a more negative score indicates the presence of more severe hypernasality. Although a comparable correlation was
found for the nasalance score of an oral text alone, the NSI 2.0 identifies patients with hypernasality more accurately compared to a single parameter approach. Although 39% of the variance in the NSI 2.0 scores could be explained by the amount of perceived hypernasality in which no significant influence of audible nasal airflow or speech intelligibility was withheld, further research is necessary to explore additional instrumental measurements and influencing variables which may explain the remaining variance. Low correlations were found between hypernasality ratings and VLHR and nasalance scores based on a single vowel. Therefore, further research could focus on the inclusion of VLHR based on single vowels extracted from words or sentences or the application of LTAS on connected speech samples in the NSI 2.0. As the NSI 2.0 correlates significantly with perceived hypernasality, it provides an easy-to-interpret severity score of hypernasality which will facilitate the evaluation of therapy outcomes, communication to the patient and other clinicians, and decisions for treatment planning, based on a multiparametric approach. However, further research is still necessary to explore the instrumental correlates of perceived hypernasality.
References


Appendix

# ===========================
# NASALITY SEVERITY INDEX 2.0
# ===========================

# ----------------------------------
# FILL-IN FORM - NASALANCE VALUES
# ----------------------------------

form NSI - NASALITY SEVERITY INDEX 2.0

    comment Please, fill in the nasalance value for the following parameters obtained by the Nasometer:
    comment To receive the proper nasalance value for the vowel /u/, comment please let the patient repeat the vowel /u/ for four times during one recording using the Nasometer.
    comment
    comment positive oral_text
    comment positive sound_u
    comment
    comment Click 'OK' to proceed.
    comment Make sure that you use a sound object of a medial segment of a sustained vowel /i/,
    comment extracted using a Hamming window (duration 0.5s), to calculate the VLHR.
    comment
    comment Script credits: K. Bettens, Y. Maryn and P. Corthals
endform

oraltext = oral_text
soundu = sound_u

# ----------------------------------
# ACOUSTIC MEASUREMENTS IN PRAAT
# ----------------------------------

#VoiceLowtonetoHightoneRatio
name$ = selected$ ("Sound")
#do ("To Pitch...", 0, 75, 600)
To Pitch... 0.05 75 600

fnul = do ("Get quantile...", 0, 0, 0.5, "Hertz")
select Sound 'name$

do ("To Spectrum...", "no")
name$ = selected$ ("Spectrum")
binWidth = Get bin width
approximateDuration = 1 / binWidth
binNum=Get number of bins
binW=Get bin width
freLow=65
freHigh=8000
altCutoff= 4.47*fnul

aFc_div_binW=altCutoff/binW
aLFP=0
aHFP=0
for i from freLow/binW+1 to freHigh/binW+1

    rV=Get real value in bin... 'i'
    iV=Get imaginary value in bin... 'i'
    mYa=(rV^2+iV^2)
    if i <= aFc_div_binW
        aLFP=aLFP+mYa
    endif
    if i > aFc_div_binW
        aHFP=aHFP+mYa
    endif

endfor
iVLHR = 10*log10(aLFP/aHFP)

# ----------------
# NSI 2.0 EQUATION
# ----------------

nsi = (13.20-(0.0824*soundu)-(0.260*oraltext)-(0.242*iVLHR))

# ---------------------------
# DRAWING PARAMETERS, NSI 2.0
# ---------------------------

# TITLE
Erase all
24
Helvetica
White
Line width... 1
Select inner viewport... 0.8 8.2 0.4 1.6
Axes... 0 1 0 1
Paint rectangle... black 0 1 0 1
Draw inner box

Viewport... 1 8 0.5 1.5
Viewport text... Centre Top 0 ##NASALITY SEVERITY INDEX 2.0##
Viewport... 1 8 0.5 1.5
Viewport text... Centre Bottom 0 ##(NSI 2.0)##

# PARAMETERS
Relationship between NSI 2.0 and perceptual judgments of hypernasality

# NSI 2.0

24
White
Select inner viewport... 0.8 8.2 3.4 4
Axes... 0 1 0 1
Paint rectangle... black 0 1 0 1
Draw inner box

Viewport... 1 8 3.5 4.5
Viewport text... Centre Top 0 NSI 2.0: ##'nsi:2'#

# GRAPHIC DISPLAY

10
Black
Select inner viewport... 1 8 4.5 5.5
Axes... -25 15 0 1
Paint rectangle... red -20 0 0 1
Paint rectangle... lime 0 10 0 1
Line width... 2
Draw rectangle... -20 15 0 1
Line width... 1
One mark top... -20 no yes no -20
One mark top... -15 no yes no -15
One mark top... -10 no yes no -10
One mark top... -5 no yes no -5
One mark top... 0 no yes no 0
One mark top... 5 no yes no 5
One mark top... 10 no yes no 10
One mark bottom... nsi no yes no ##'NSI 2.0##
Line width... 2
One mark bottom... nsi no no yes NSI 2.0

# REPLACE SELECTION WINDOW

Select inner viewport... 0.8 8.2 0.4 5.7
Copy to clipboard
Select inner viewport... 0.8 8.2 7 8
CHAPTER 8

SHORT-TERM EFFECT OF SHORT, INTENSIVE SPEECH THERAPY ON ARTICULATION AND RESONANCE IN UGANDAN PATIENTS WITH CLEFT (LIP AND) PALATE


Abstract

**Objective.** The purpose of the current study was to assess the short-term effectiveness of short and intensive speech therapy provided to patients born with cleft (lip and) palate (C(L)P) in terms of articulation and resonance.

**Methods.** Five Ugandan patients (age 7-19 years) born with non-syndromic C(L)P received six hours of individualized speech therapy in three or four days. Speech therapy focused on correct phonetic placement and contrasts between oral and nasal airflow and resonance. Speech evaluations performed before and immediately after speech therapy, including perceptual and instrumental assessment techniques, were compared.

**Results.** Post-therapy, improvement of speech was noted for most of the patients, although to varying degrees. Clinically relevant progress of objective nasalance values and/or articulation was obtained in four patients. Overall, two patients showed normal speech intelligibility, while three patients required additional speech therapy.

**Conclusion.** These preliminary short-term results demonstrate that short and intensive speech therapy can be effective for patients with C(L)P in countries with limited access to speech-language therapy. However, further research is needed on the long-term effectiveness and the advantages of applying this treatment protocol in countries with good access to speech therapy.

**Key words.** Cleft palate; speech therapy; effectiveness; articulation; resonance; airflow

**Learning outcomes.** The reader will be able to (1) list the challenges in resource poor-countries to achieve access to speech-language therapy services, (2) describe when the application of speech therapy is appropriate in patients with C(L)P, (3) describe the speech therapy that can be applied to reduce compensatory articulation and resonance disorders in patients with C(L)P, and (4) list the (possible) advantages of short, intensive speech therapy for both resource-poor and developed countries.
Introduction

In children born with cleft (lip and) palate (C(L)P), resonance and articulation disorders are often observed as a result of structural deviations of the sound production mechanism. Even after palatal closure, compensatory articulation and resonance disorders may persist, despite advances in surgical treatment of congenital orofacial clefts (Hardin-Jones & Jones, 2005). As these speech disorders frequently affect speech intelligibility, intervention is often required. To select the best treatment option (i.e. secondary surgery, speech therapy or prosthetic intervention), the cause of the velopharyngeal dysfunction and related resonance and articulation errors has to be defined. In the case of velopharyngeal insufficiency (VPI), adequate closure of the velopharyngeal mechanism cannot be obtained due to an anatomic defect such as a short or malfunctioning velum, which may be observed after cleft palate repair (Kummer, 2011). This can result in hypernasality and audible nasal emission and/or turbulence due to excessive nasal airflow during the production of oral sounds. Consequently, obligatory articulation errors, such as nasalization of oral consonants and decreased intraoral pressure during the production of pressure consonants, and compensatory articulation errors, such as glottal stop substitutions, pharyngeal and nasal fricatives, may arise. As the deficit originates in an abnormal anatomy, speech therapy alone cannot resolve these speech disorders, resulting in the need for additional surgery or prosthetic intervention (Kummer, 2011). Velopharyngeal mislearning, on the other hand, refers to an abnormal articulation despite normal anatomic structures and physiology of the velopharyngeal mechanism. This abnormal articulation can result from persisting compensatory speech errors after correction of the velopharyngeal structures and function. As adequate closure of the velopharyngeal valve is possible, speech therapy is appropriate to correct these articulatory deficits (Kummer et al., 2015).

Although the variability in presence, type and severity of speech disorders is huge in patients with C(L)P, the majority of these patients will still need speech therapy after palatal closure or secondary surgery (Hardin-Jones & Jones, 2005). In children with mild to moderate speech impairment, speech therapy is generally provided during several years with a frequency of two sessions of 21 to 30 minutes per week (Mullen & Schooling, 2010). Even though this approach might be effective (Pamplona et al., 1999; Pamplona et al., 2005), it is burdensome for patients and their parents as well as expensive for health insurance. Moreover, for patients living in remote areas or in countries with limited speech therapy facilities, this conventional speech therapy approach might be unfeasible. In Uganda, for example, 19 speech-language pathologists (SLPs) were active in 2013, including 4 long-term SLP expatriates as well as 15 graduates from the bachelor program for Speech-Language Therapy at the Makarere University, Kampala, Uganda, resulting in the availability of 0.6 SLPs per 1.000.000 citizens (Luyten, 2014). Given that the incidence of C(L)P in Uganda lies between 0.73 (Dreise et al., 2011) and 1.34 (Kalanzi et al., 2013) per 1000 live births and the annual number of births in Uganda is estimated at 1.5 million (UNICEF, 2013), a total of 1128 to 2070 children are born with C(L)P each year. Although not all of them will require speech
therapy, the availability of only 19 SLPs in this country implies long travel distances for most patients so that the traditional models for delivering speech therapy to children with C(L)P are not adequate to reach all patients in need (D'Antonio & Nagarajan, 2003). In order to counter these limitations, short and intensive speech intervention with or without overnight stay might be a solution (Scherer, 2014).

Previous studies regarding the effectiveness of the speech camp model, i.e. short and intensive speech intervention, overall led to encouraging results concerning articulation (Pamplona et al., 2005; Prathanee et al., 2011; Van Demark & Hardin, 1986). However, little is known about the effect of such speech therapy protocol on resonance disorders. Van Demark and Hardin (1986) reported a significant increase in the percentage of correctly produced sounds between pre- and post-therapy examinations when patients with repaired cleft (lip and) palate received four hours of speech therapy daily during 26 days. Nevertheless, no significant improvement in hypernasality severity ratings was found. Pamplona et al. (2005) observed a comparable significant decrease in severity of compensatory articulation errors after attending either a speech camp of three weeks (4h therapy/day during 5 days/week), or receiving one hour speech therapy twice a week for a period of twelve months. Moreover, Prathanee et al. (2011) observed a significant reduction in the number of articulation errors after a four-day speech camp (18h therapy) as well as after a one-day follow-up session (6h therapy) six months later.

The need for new models of speech therapy schedules for patients living in remote areas or in countries with limited speech therapy facilities as well as the above-mentioned encouraging results of short and intensive speech intervention led to the establishment of a similar speech therapy project in the Comprehensive Rehabilitation Services in Uganda (CoRSU) hospital, a non-profit and non-government organization specialized in surgical treatment of orofacial clefts. In the specialist plastic and reconstructive unit of CoRSU, one permanent plastic and reconstructive surgeon (A.H.) treats all patients with C(L)P, thanks to financial support by Smile Train. However, only limited attention could be given so far to the inclusion of speech therapy for these patients due to the limited availability of SLPs in Uganda. Until recently, patients who needed speech therapy were referred to Mulago Hospital in the capital city. However, long travel distances and lack of means prevented most patients from receiving proper assistance. Therefore, this study took a first step in creating the possibility to provide speech therapy in CoRSU. During this project, six hours of speech therapy were provided by two foreign SLPs in three or four days to patients with articulation and resonance disorders caused by congenital C(L)P. The purpose of the current study was to verify the short-term effectiveness of such short, intensive speech therapy on the articulation and resonance in five patients born with C(L)P.
Method

Participants

A database including all patients with C(L)P (n=199) who were seen for speech assessment between January 2011 and May 2013 during the VLIR-UOS project of our research unit (ZEIN2009EL28) was used to select the participants for this study. Criteria for selection and inclusion were (1) repaired hard and soft palate, (2) at least one orofacial surgical treatment undergone in CoRSU, Kisubi, Uganda, performed by dr. A. H., (3) sufficient knowledge of English according to the grade-level mentioned in the patient’s file, (3) full speech assessment performed with good cooperation according to age at a previous consultation, (4) attended the follow-up assessments during the VLIR-UOS project, and (5) presence of remarkable articulation errors and resonance disorders that influence speech intelligibility based on previous speech assessments performed by our research unit. Patients with a syndromic cleft (based on clinical examination by the treating surgeon (A. H.) as genetic testing was unavailable) were excluded. Although the cause of resonance and articulation disorders should be known before the start of speech therapy, the presence of VPI was not included in the exclusion criteria due to the unavailability of the equipment to perform nasoendoscopic or videofluoroscopic assessments. Twelve patients fulfilled the above-mentioned criteria. Due to organizational limitations, seven patients were randomly selected to participate in the study. The patients or their relatives were contacted via phone by the social worker of CoRSU, who provided information about the purpose of the study and invited them to participate voluntarily. Six of the seven patients joined the study, although one patient terminated his participation early on citing personal reasons. All included patients passed a supralinearhearing screening (see assessment procedures). This resulted in the inclusion of five Ugandan patients with non-syndromic orofacial clefts in this speech therapy project. Specific patient information is provided in Table 8.1. This research was approved by the Ethical Committee of the Ghent University Hospital, Belgium (EC2011/269). All participants and their legal representatives, if they were not yet 18, were informed about the study, both orally and by letter. Informed consent was signed by the participant or his/her legal representative.
Table 8.1 Demographic, cleft, and surgical details for the participating patients.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Gender</th>
<th>Age</th>
<th>Cleft type</th>
<th>Age at lip closure</th>
<th>Age at palatal closure</th>
<th>Palatal closure at CoRSU (performed by dr. A. H.)</th>
<th>Secondary surgery (performed at CoRSU by dr. A. H.)</th>
<th>Oronasal fistula</th>
<th>Previous speech therapy</th>
<th>Grade-level</th>
<th>Work</th>
<th>Employment mother</th>
<th>Employment father</th>
<th>Mother language</th>
<th>Use of English outside school</th>
<th>Literacy</th>
<th>Grade</th>
<th>Employment father</th>
<th>Use of English outside school</th>
<th>Literacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>female</td>
<td>10;8 years</td>
<td>UCLP (right)</td>
<td>2 months</td>
<td>11 months</td>
<td>yes</td>
<td>- speech improving surgery using buccal flap*(8;5 years)</td>
<td>no</td>
<td>12 hours</td>
<td>grade 6 at primary school (attends every day)</td>
<td>n.a.</td>
<td>NGO coworker</td>
<td>n.a.</td>
<td>English</td>
<td>yes, everywhere</td>
<td>literate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>female</td>
<td>17;5 years</td>
<td>CP</td>
<td>n.a.</td>
<td>6 months</td>
<td>no</td>
<td>- fistula repair (14;0 years)</td>
<td>no</td>
<td>0 hours</td>
<td>grade 5 at primary school (attends mostly)</td>
<td>n.a.</td>
<td>cleaning lady</td>
<td>driver</td>
<td>English</td>
<td>yes, with friends</td>
<td>literate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>male</td>
<td>15;2 years</td>
<td>CP</td>
<td>n.a.</td>
<td>13;5 years</td>
<td>yes</td>
<td>- speech improving surgery using buccal flap*(14;6 years)</td>
<td>no</td>
<td>no</td>
<td>until grade 7 at primary school (attended sometimes)</td>
<td>n.a.</td>
<td>driver</td>
<td>n.a.</td>
<td>Rutooro</td>
<td>no</td>
<td>Luganda</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>female</td>
<td>19;7 years</td>
<td>UCLP (right)</td>
<td>unknown</td>
<td>3 years</td>
<td>no</td>
<td>- fistula repair (18;3 years)</td>
<td>no</td>
<td>no</td>
<td>until grade 4 at secondary school (attended every day)</td>
<td>n.a.</td>
<td>gas station attendant</td>
<td>n.a.</td>
<td>Luganda</td>
<td>no</td>
<td>yes, with friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>female</td>
<td>7;4 years</td>
<td>UCLP (left)</td>
<td>2 months</td>
<td>2 months</td>
<td>yes</td>
<td>- nasal correction (4;1 years)</td>
<td>yes</td>
<td>no</td>
<td>grade 2 at primary school (attends every day)</td>
<td>n.a.</td>
<td>graphic designer</td>
<td>n.a.</td>
<td>Kakwa</td>
<td>yes, with friends</td>
<td>literate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

UCLP: unilateral cleft lip and palate, CP: cleft soft and hard palate, n.a.: not applicable
*Mann et al. (2011)
**Assessment procedures**

**Hearing assessments.** All assessments were carried out in English in a clinic room at CoRSU. The patients’ hearing was screened, using the Rinne and Weber tuning fork tests (Turner, 1990) as well as a supraliminar hearing screening. For the Rinne test, an activated tuning fork was applied to the mastoid bone and was, subsequently, held in front of the ear. A positive result was achieved when the sound was heard louder in front of the ear compared with behind the ear. During the Weber test, a tuning fork was activated and placed on the center of the forehead. A normal response implied no lateralization of the sound energy to either ear. Additionally, monosyllabic digits between one and twelve were presented with decreasing signal-to-noise ratios (SNRs). The noise-mixed digits were presented through a standard available speaker (PS speaker, sp-690), calibrated to fix the noise level at 65dB SPL and placed at a distance of 90cm in front of the patient. The noise-mixed digits were presented one-by-one starting at an SNR of 6dB, after which the patient was asked to repeat the digit heard. A series of six digits was presented using the same SNR. If the patient was able to repeat at least 50% of the digits presented using the same SNR correctly, the SNR was decreased by one until the patient failed to repeat 50% of the presented digits correctly. Hearing disorders were excluded when the patient was able to repeat at least 50% of the digits with an SNR of -5dB correctly. This cutoff was chosen based on biological calibration. Two SLPs whose hearing levels were primarily determined by tonal audiometry at the Ghent University Hospital in Belgium (PTA<15dB HL on each ear), performed the test on the same day and in the same conditions as the patients. As both SLPs were able to repeat at least 50% of the offered numbers correctly at an SNR of -5dB, this threshold was withheld as a representation of sufficient hearing. We opted for these tests because the conditions required to perform tonal audiometry were not satisfied.

The results for the individual patients’ hearing assessments are summarized in Table 8.2. Patient 2 was not able to perform the tuning fork tests, because she did not understand the instructions. Positive results were obtained for all patients regarding the Rinne tests. Hence, the Weber test revealed normal results in patient 3 and 4, while in patient 1 and 5 lateralization to the left was observed. All thresholds for the supraliminar hearing screening varied between SNRs of -5 and -6dB. Consequently, none of the hearing screenings seemed to reveal an indication for significant hearing disorders.
Table 8.2 Individual results for the hearing assessments.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Rinne Left ear</th>
<th>Rinne Right ear</th>
<th>Weber</th>
<th>Supraliminal hearing screening (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient 1</td>
<td>+</td>
<td>+</td>
<td>left</td>
<td>SNR -6</td>
</tr>
<tr>
<td>Patient 2</td>
<td>t.n.p.</td>
<td>t.n.p.</td>
<td>t.n.p.</td>
<td>SNR -6</td>
</tr>
<tr>
<td>Patient 3</td>
<td>+</td>
<td>+</td>
<td>middle</td>
<td>SNR -6</td>
</tr>
<tr>
<td>Patient 4</td>
<td>+</td>
<td>+</td>
<td>middle</td>
<td>SNR -5</td>
</tr>
<tr>
<td>Patient 5</td>
<td>+</td>
<td>t.n.p.</td>
<td>left</td>
<td>SNR -6</td>
</tr>
</tbody>
</table>

t.n.p.=testing not possible

Speech assessments. An identical battery of speech assessments was performed in the same order by the non-treating SLP prior to and after the speech therapy sessions. During these speech assessments, patients were first asked to sustain the vowel /i/ for at least 2 seconds in front of a unidirectional condenser microphone (Samson, C01U) placed at 10cm from the patient’s mouth. The vowel was audio-recorded using PRAAT software, version 5.3.78 (Boersma & Weenink, 2014) at a sampling frequency of 44100Hz. Second, a speech sample was video-recorded using a Sony HDR-UX1 camera with a high quality built-in microphone and was simultaneously audio-recorded with PRAAT software using a unidirectional condenser microphone (Samson, C01U). The speech sample consisted of the repetition of 12 oral and 3 nasal simple English sentences of the MacKay-Kummer Simplified Nasometric Assessment Procedures (SNAP) test (Kummer, 2005), which were modeled by the SLP. These sentences can be classified in five sentence groups according to the main consonants’ characteristics: bilabials, alveolars, velars, sibilants, and nasals. In addition, the patients counted from 1 to 10 and from 60 to 70 and recited the days of the week. Third, the standardized picture-naming test Photo Articulation Test – Third Edition (PAT-3) (Lippke et al., 1997) was administrated and video-recorded (Sony HDR-UX1 camera). As such, 72 high frequency English words were elicited by colored pictures (i.e. 40 monosyllabic, 28 disyllabic and 4 three-syllabic words), in which all English consonants occurred in all permissible syllable positions as well as in common consonant clusters. While most patients named the pictures, patient 3 repeated the words read out by the SLP, given his limited knowledge of English vocabulary, despite the fact that he accomplished 7 years of primary school during which all lessons were provided in English. Finally, a KayPentax Nasometer (model II 6450) (NJ, Lincoln Park) was used to obtain objective nasalance values. Nasalance values were collected for the sustained vowel /u/, for the five sentence groups of the SNAP test (Kummer, 2005), as well as for the oronasal rainbow passage and the oral zoo passage, which were elicited via phrase repetition in all patients. The Nasometer was calibrated according to the manual’s instructions. Subsequently, the sound separator plate of the headgear was placed beneath the nostrils. The oral and nasal acoustic energy collected by the microphones in front of the mouth and the nose were processed by a portable computer. As such, a nasalance score (%) was calculated by dividing the nasal acoustic energy by the oral-plus-nasal-acoustic-energy and multiplying the quotient with 100. According to Watterson and Lewis (2006), a difference of at least 10% between nasalance
scores obtained from the same stimulus (i.e. sentences and text passages) before and after therapy was considered to be a genuine change. Additionally, in accordance with Bettens et al. (2016), the nasalance scores for the vowel /u/ (%) and the oral zoo passage (%) as well as the voice low tone to high tone ratio (VLHR) of the vowel /i/ with a cutoff score of 4.47*F0Hz were used to calculate the Nasality Severity Index 2.0 (NSI 2.0) following the equation NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%)) – (0.260 x nasalance oral text (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)). VLHR was achieved from the audio-recordings of the vowel /i/, as described by Lee et al. (2003). After a Fast Fourier Transformation, a sample of 0.5s was analyzed using a Hamming window. Based on a cutoff frequency of 4.47*F0Hz, the spectrum was divided into a low frequency (65Hz to cutoff) and high frequency spectrum (cutoff to 8000Hz). After summation of the power of each frequency component of the spectra, VLHR was determined by the ratio of the low frequency to the high frequency spectrum, expressed in dB. This multiparametric index identifies the presence of hypernasality more accurately than a single parameter approach (Bettens et al., 2016). Hypernasality is absent when the resulting value is positive; hypernasality is present when a negative value is obtained.

**Perceptual evaluation.** The audio-recorded speech samples consisting of sentence repetition and automatic speech (i.e. counting and reciting the days of the week) were anonymized and randomized. Perceptual evaluations were carried out independently by the investigating SLP (also the principal investigator of this study), the treating SLP and another non-Ugandan Dutch-speaking Flemish SLP who was blind for the purposes of the study. The audio-recordings were played back to the SLPs via headphones. The definitions and rating system of the Cleft Audit Protocol for Speech – Augmented (CAPS-A) (John et al., 2006) were applied for perceptual evaluation of hypernasality (absent, borderline, mild, moderate, severe), hyponasality (absent, mild, marked), audible nasal emission and nasal turbulence (absent, occasionally, frequently), voice disorder (absent, disturbed voice quality) as well as speech intelligibility (i.e. the ability to understand the speech that is heard) (normal; different from other children’s speech, but not enough to cause comment; different enough to provoke comment, but possible to understand; only just intelligible to strangers; impossible to understand). In addition, the absence or presence of cul-de-sac resonance was evaluated. The interjudge reliability between SLPs was calculated for each variable by the ratio of identical to total judgments. The overall interjudge reliability varied from 81% to 84%. Ranges for the specific interjudge reliabilities were 60%-80% for hypernasality, 80%-100% for hyponasality and nasal turbulence, 60%-70% for audible nasal emission and 60%-90% for speech intelligibility. Regarding voice disorder and cul-de-sac resonance, all interjudge reliability scores reached 100%. For all variables, the patients’ scores were determined by the median score of the three SLPs.

**Articulation analyses.** The videotaped speech samples of the picture-naming test PAT-3 (Lippke et al., 1997) and the 15 sentences of the SNAP test (MacKay & Kummer, 1994) were played back via headphones and were phonetically transcribed by the investigating SLP
using the International Phonetic Alphabet (IPA) (International Phonetic Association, 1999), the IPA extensions (Duckworth et al., 1990) as well as additional symbols to describe specific cleft-related articulation errors (Peterson-Falzone et al., 2006). Moreover, 20% of the words (15/72) and sentences (3/15) were transcribed by the independent SLP. The narrow consonant-for-consonant interrater reliability was 69%. Further analyses were based on the 18 consonants (i.e. /p/, /b/, /t/, /d/, /k/, /ɡ/, /s/, /z/, /ʃ/, /ʒ/, /l/, /r/, /n/, /m/, and /ŋ/) found in English as well as in the patients’ African mother language. As experience showed that many Ugandans without clefts show difficulties with the correct pronunciation of the English fricatives /θ, ð, ʃ, ʒ/ and affricates /tʃ, dʒ/ (Luyten et al., 2014), these speech sounds were excluded from the analysis. Moreover, it decreased the influence of the listeners’ language background (Hutters & Henningsson, 2004), considering that only consonants identical to Dutch consonants were analyzed.

A phonetic analysis was performed at the segmental level by comparing the consonants with the target speech sounds. At the word and sentence level, the overall percentage of correct consonants was calculated as well as the percentage of correctly produced plosives and fricatives. A correct production included a correct place and manner of articulation as well as correct airflow direction. Furthermore, the occurrence frequency (%) of all specific phonetic errors was determined by the ratio of actual to potential occurrences.

**Speech therapy**

Speech therapy was provided in March or September 2014 in a clinic room of CoRSU by two non-Ugandan SLPs (one SLP per patient). Treatment was implemented in English as the patients spoke different native languages. English is one of the official languages of the country, resulting in a deep embedding of this language in education, media, religion, economics and politics (Mpuga, 2003). According to their grade-level and a short conversation, all patients showed sufficient knowledge of English. Moreover, both SLPs fluently spoke and comprehended the English language, although they were native speakers of Dutch. During three or four consecutive days, each patient received six hours of speech therapy (1 hour per session). Additionally, a short revision session of ten minutes was organized shortly before the post-therapy speech assessment.

As the SLPs focused on the consonants that significantly influenced speech intelligibility, the target consonants differed between patients. If several speech sounds affected the speech intelligibility in the same way, those sounds that are acquired first during normal speech development were firstly addressed. Individual phonetic articulation therapy was provided (Van Riper, 1978) in combination with principles of motor learning (Maas et al., 2008; Schmidt & Lee, 2005). Phonetic therapy was chosen because of the older age of the participants. Principles of motor learning were applied to diminish the motor speech program of the compensatory articulation and to establish new motor routines for these speech sounds. In a first phase, prepractice considerations were ensured (Maas et al., 2008; Schmidt & Lee, 2005). First, the motivation of the patient was re-enforced by providing
information about the relevance of the practice tasks, more specifically to improve speech intelligibility. Second, the identification of the selected speech sound by observing the characteristics of the specific consonant was established. The correct place of articulation was showed on an anatomic model, after which the SLP showed the correct place of articulation by herself and pointed at the speech articulators of the patient. Awareness of correct and error productions was initiated by providing examples of the correct speech sound using a correct articulatory placement and the patient’s sound production produced by the SLP with specific attention to the resonance and airflow deviation. To support the identification of nasal resonance and airflow, pictures of a nose and mouth were provided. Additionally, phonological methods were applied to support the establishment of the contrast between oral and nasal resonance and nasal airflow in consonants, more specifically using the distinctive feature approach (Costello & Onstine, 1976; McReynolds & Bennett, 1972). In a next step, auditory discrimination of correct and error consonant productions produced by the SLP was trained. After this, the patient was encouraged to compare his/her own productions with those of the SLP to detect the differences. When these auditory discrimination skills were achieved, the practice phase was started by eliciting a correct sound production by using techniques of progressive approximation, direct auditory stimulation, phonetic placement, modification of another known speech sound and/or pinching the nose. The use of auditory, visual and tactile feedback was emphasized during the whole therapy. Additionally, augmentative feedback (Maas et al., 2008) about the sound production (i.e. articulatory placement and acoustic signal) was provided by the SLP either immediately or delayed, so that the patient could evaluate his/her production first by him/herself. Once a correct sound production could be evoked, stabilization of the correct production was attained by variable practice, more specifically by repeating, prolonging and varying loudness levels. Depending on the progress of the patient, therapy focused successively on the production of the target speech sound(s) in syllables, words, sentences, texts and transfer to spontaneous speech. A next level was introduced by the SLP when the patient was able to correctly produce the target speech sound in approximately 90% of the time with minimal cues from the SLP. When many errors were observed in the next level, the patient was encouraged to take a step back to the former level. For example, when a patient had difficulties producing the target speech sound correctly in a particular word, he/she was asked to produce the sound in isolation until a correct production was observed, after which the word was asked again. This was done in order to review the correct motor speech schema and create a positive therapy environment that praised articulatory successes in order to maintain the patient’s motivation.
Results

Speech assessments

The individual results of the pre- and post-therapy speech assessments are provided in Table 8.3.

Patient 1

Prior to speech therapy, patient 1 presented with severe hypernasality and frequent audible nasal emission. This was reflected in high nasalance values for the oral sentences, the oral text and the oronasal text as well as in a negative NSI. Articulation was mainly characterized by substitution of /s,z/ by a nasal fricative (words: 90%, sentences: 95%), weak production of plosives /p,b,t,d,k,g/ (words: 41%, sentences: 26%) and the production of /t,d/ followed by a short nasal fricative (words: 38%, sentences: 76%). Additionally, (inter-) dental production of the apico-alveolar consonants /t,d,l,n/ was observed (words: 44%, sentences: 0%). Overall, speech intelligibility was judged to be mildly impaired (i.e. different enough to provoke comment, but possible to understand most speech). Although the presence of hypernasality, audible nasal emission and weak productions of plosives are normally classified as obligatory resonance and articulation errors, this patient was occasionally able to produce plosives with enough intraoral air pressure, suggesting the possibility to close the velopharyngeal mechanism properly and justifying the application of speech therapy.

As the compensatory articulation of the sounds /s/ and /z/ was mostly disturbing speech intelligibility, the therapy concentrated on the reduction of this compensatory articulation by evoking the correct production of the consonants /s/ and /z/ in isolation as well as in syllables, words, sentences, texts and spontaneous speech. Additionally, the importance of oral air flow was particularly emphasized during the production of pressure consonants (i.e. plosives, fricatives and affricates), by using bio- and augmentative feedback, to optimize the production of /s/ and /z/ and to reduce the weak production of plosives, the amount of hypernasality and audible nasal emission.

Post-therapy, neither resonance disorders, nor airflow deviation errors were perceptually observed. This was confirmed by the instrumental assessments, resulting in nasalance scores within normal limits compared to the normative data (Luyten et al., 2012) and a positive NSI. The consonants /s,z/ were produced either correctly (words: 48%, sentences: 14%), with a strident (words: 48%, sentences: 86%) or were substituted by /t/ followed by a short /s/ sound (words: 3%, sentences: 0%). The occurrence frequency of weak plosives (words: 6%, sentences: 0%) and production of /t,d/ followed by a short nasal fricative (words: 10%, sentences: 0%) decreased strongly, while (inter-) dentalization of alveolar consonants /t,d,l,n/ persisted (words: 52%, sentences: 32%). Overall, speech intelligibility was judged to be normal.
**Patient 2**

Prior to speech therapy, patient 2’s speech was characterized by severe hypernasality, occasional audible nasal emission and frequent nasal turbulence. In accordance, increased nasalance values were observed for the oral sentences and the oronasal and oral text compared to the reference values (Luyten et al., 2012). Moreover, a strongly negative NSI value was obtained. Regarding articulation, the consonants /s,z/ and /k,g/ were mostly substituted by a pharyngeal fricative (words: 93%, sentences: 91%), or a pharyngeal plosive (words: 7%, sentences: 44%), respectively. Overall, speech intelligibility was categorized as moderately disturbed (i.e. only just intelligible to strangers). With 72% of all plosives /p,b,t,d,k,g/ being correctly produced at the word level and only 30% at the sentence level pre-therapy, the presence of pharyngeal fricatives and plosives was assumed to be behavioral in origin and therefore treatable with articulation therapy.

Speech therapy focused on the correct production of the consonant /s/ in isolation as well as in syllables, words and sentences, with special attention to the correct direction of airflow.

After speech therapy, similar resonance disorders and airflow deviation errors were still present based on the perceptual evaluations as well as the objective NSI measurements. However, a slight, but genuine decrease (i.e. ≥10% (Watterson & Lewis, 2006)) was noticed for the nasalance values of bilabial, velar and sibilant sentences. Furthermore, articulation analyses revealed a remarkable decrease in occurrence frequency of pharyngeal fricatives as a substitute for /s,z/ (words: 67%, sentences: 74%). Other observed articulation errors for these sounds included omissions at the word level (3%) and substitution by /f,k/, and decreased frication at the sentence level (18% and 4%, resp.). To the contrary, a decrease was noticed in the percentage of correctly produced plosives at the word level. This was due to an increase in the occurrence frequency of pharyngeal plosives which were used as a substitute for the sounds /k,g/ (50%). Speech intelligibility was still overall judged to be moderately impaired (i.e. only just intelligible to strangers).

**Patient 3**

Prior to speech therapy, patient 3 presented with severe hypernasality, which resulted in high nasalance values for the oral sentences and oronasal and oral texts compared to the reference values (Luyten et al., 2012), as well as in a strongly negative NSI score. Articulation was characterized by a substitution of glottal stops for /t,d/ (words: 43%, sentences: 53%), /k,g/ (words: 73%, sentences: 88%) and /s,z/ (words: 13%, sentences: 54%). Moreover, the consonants /s,z/ were frequently omitted (words: 74%, sentences: 38%) and weak production of plosives /p,b,t,d,k,g/ occurred (words: 38%, sentences: 20%). Overall, speech intelligibility was judged to be severely disturbed (i.e. impossible to understand). As 50% of the plosives /p,b/ were correctly produced at the word level pre-therapy, the velopharyngeal mechanism was assumed to close properly. Additionally, the
production of glottal stops was considered to be a compensatory articulation error which could be reduced by speech therapy.

Speech therapy particularly focused on the correct production of /t/ in isolation and in syllables, words and sentences. In addition, the correct production of /s/ was targeted starting from /t/ in isolation as well as in syllables and words. The importance of oral air flow was emphasized during the production of these pressure consonants by using repetitive drill and augmentative feedback.

After speech therapy, hypernasality decreased to a moderate level. In contrast, only the nasalance scores of the velar sentences showed a genuine decrease of 12% and the NSI did not change. Regarding articulation, the consonants /t,d/ were either produced correctly (words: 43%, sentences: 59%) or followed by a short /s/ sound (words: 33%, sentences: 6%). Moreover, the frequency of correctly produced /k,ɡ/ consonants increased remarkably (words: 65%, sentences: 28%) and the occurrence frequency of weakly produced plosives decreased (words: 8%, sentences: 7%). While in sentences, the consonants /s,z/ were still omitted (32%) or substituted by a glottal stop (60%), the number of correct productions of /s,z/ at the word level increased (39%). However, substitution of /s,z/ by /t/ (32%) and the addition of a consonant /t/ prior to the production of /s/ (26%) at the word level was remarkable. Despite the advances in resonance and articulation, speech intelligibility was still judged to be severely disturbed.

**Patient 4**

Prior to speech therapy, patient 4’s speech was characterized by severe hypernasality without audible nasal emission or turbulence. In accordance, moderately increased nasalance values were observed for all oral and oronasal speech samples compared to the reference values (Luyten et al., 2012) and a negative NSI score was obtained. Regarding articulation, the consonants /s,z/ were frequently omitted (words: 53%, sentences: 41%) or substituted by a weakly produced /t/ sound (words: 17%, sentences: 41%). Furthermore, substitution of /k,ɡ/ by an often weakly produced alveolar or palatal stop frequently occurred (words: 73%, sentences: 72%). Finally, weak productions of the consonants /p,b/ (words: 20%, sentences: 73%), /t,d/ (words: 52%, sentences: 71%) and /f,v/ (words: 80%, sentences: n.a.) were noticed. Overall, speech intelligibility was judged to be severely disturbed (i.e. impossible to understand). Although the presence of hypernasality and weak productions of plosives are normally classified as obligatory resonance and articulation errors, this patient was able to produce 80% of the plosives /p,b/ with enough intraoral air pressure at the word level pre-therapy, suggesting the possibility to close the velopharyngeal mechanism properly. Moreover, a re-repair of the palate using the Furlow technique was performed recently to close the soft palate and to advance velopharyngeal closure. The presence of hypernasality and articulation errors could therefore be persistent despite an assumed functioning of the velopharyngeal mechanism which justified the application of speech therapy in this patient.
Speech therapy focused on the buildup of more intraoral pressure during the production of /t/ in syllables, words and sentences. In addition, the correct production of /s,z/ was learned starting from the articulation place of the sound /t/, in isolation as well as in syllables, words and sentences with the nose closed.

The post-therapy perceptual evaluations revealed severe hypernasality and frequent audible nasal emission, which was reflected in higher objective nasalance values for all speech samples and a worse NSI score compared with the pre-therapy condition. Regarding articulation, the consonants /s,z/ were produced with a correct place and manner of articulation or with (inter-) dentalization, albeit the productions were always accompanied with severe audible nasal emission (words: 90%, sentences: 64%). Moreover, presence of audible nasal emission led to increased occurrence frequencies for weak production of /p,b/ (words: 72%, sentences: 91%), /t,d/ (words: 81%, sentences: 56%) and /f,v/ (words: 100%, sentences: n.a.). Similar occurrence frequencies were observed for substitution of /k,g/ by an often weakly produced alveolar or palatal stop (words: 69%, sentences 61%) compared to the pre-therapy assessment. Finally, speech intelligibility was still categorized as severely disturbed.

**Patient 5**

Prior to speech therapy, patient 5 presented with normal resonance, but occasional audible nasal emission. The nasalance scores of the sibilant sentences and the oronasal and oral text were high in comparison with reference values (Luyten et al., 2012) and a negative NSI was achieved. Articulation analyses showed a high occurrence frequency for /s,z/ produced as a nasal fricative (words: 97%, sentences: 95%) as well as for (inter-) dental productions of the consonants /t,d/ (words: 52%, sentences: 65%). Overall, mildly disturbed speech intelligibility was observed (i.e. different enough to provoke comment, but possible to understand most speech). As normal resonance was observed in combination with only one manifest compensatory articulation error (production of a nasal fricative for /s,z/), this articulation error was assumed to be behavioral in origin and therefore treatable with articulation therapy.

Speech therapy focused on the correct production of the consonants /s,z/ in isolation as well as in syllables, words, sentences, texts and transfer to spontaneous speech. The importance of oral airflow was particularly emphasized.

Post-therapy, no resonance disorders or airflow deviation errors occurred, which was objectively confirmed by nasalance values lying within normal limits (Luyten et al., 2012) for all oral and oronasal speech samples and a positive NSI score. Regarding articulation, the consonants /s,z/ were either produced correctly (words: 67%, sentences: 68%), dental (words: 3%, sentences: 0%), or with decreased friction (words: 30%, sentences: 32%). However, the (inter-) dental production of the consonants /t,d/ remained (words: 57%,...
sentences: 65%). Consequently, speech intelligibility changed from mildly disturbed prior to speech therapy to normal post-therapy.

Discussion

For most Ugandan patients with C(L)P, the limited availability of SLPs and the long travel distances hamper the chance to receive speech therapy on a regular basis (e.g. 2 x 30 minutes per week). Therefore, the traditional models for delivering speech therapy were replaced by short, intensive speech therapy in the Speech-Language Therapy Department of CoRSU. Consequently, the main purpose of the current study was to verify the short-term effectiveness of six hours speech therapy delivered in three or four days to patients with speech disorders caused by a history of C(L)P.

Similar to previous studies regarding the effectiveness of short and intensive speech intervention (Pamplona et al., 2005; Prathanee et al., 2011; Van Demark & Hardin, 1986), short-term improvement of speech was noted for most of the patients, although to varying degrees. Three out of five patients (patients 1, 2, and 5) showed a genuine decrease of the objective nasalance values of the sentences and/or text passages (i.e. ≥10% (Watterson & Lewis, 2006)) and four patients (patients 1, 2, 3, and 5) showed considerable improvements in the percentage of correct productions of the treated consonants, particularly at the word level. Moreover, perceptual evaluation of hypernasality (patients 1 and 3), audible nasal emission (patients 1 and 5) and speech intelligibility (patients 1 and 5) each improved in two out of five patients. Despite the genuine decrease in nasalance values, the clinical relevance of this decrease can be doubted in patient 2 considering the similar degree of perceived hypernasality before and after speech therapy. Patient 3 on the other hand showed no genuine decrease in nasalance scores, but was judged to have a lower degree of hypernasality after therapy. This contradiction highlights the importance of combining instrumental measurements and perceptual judgments in the assessment of resonance disorders. Regarding patient 4, the perceptual evaluation of audible nasal emission and objective nasalance measurements deteriorated after speech therapy. This can be explained by a change in the articulation abilities, given that the sounds /s,z/ were no longer omitted or substituted by /t/ (i.e. stopping), but were produced with a correct place and manner of articulation, albeit accompanied with severe audible nasal emission. The resonance and articulation errors of this patient pre-therapy were assumed to be the result of velopharyngeal mislearning as she was able to produce 80% of the plosives /p,b/ with sufficient intraoral air pressure at the word level and a re-repair of the soft palate was already performed. Therefore, it was expected that speech therapy could reduce the observed resonance and articulation errors. However, post-therapy results indicated the presence of persistent velopharyngeal insufficiency, which could unfortunately not be verified as the equipment to perform nasoendoscopic or videofluoroscopic assessments was not available.
Table 8.3 Individual results for the pre- and post-therapy speech assessments.

<table>
<thead>
<tr>
<th>Perceptual evaluation</th>
<th>Patient 1 Pre</th>
<th>Patient 1 Post</th>
<th>Patient 2 Pre</th>
<th>Patient 2 Post</th>
<th>Patient 3 Pre</th>
<th>Patient 3 Post</th>
<th>Patient 4 Pre</th>
<th>Patient 4 Post</th>
<th>Patient 5 Pre</th>
<th>Patient 5 Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypernasality</td>
<td>severe</td>
<td>absent</td>
<td>severe</td>
<td>severe</td>
<td>severe</td>
<td>moderate</td>
<td>severe</td>
<td>severe</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>Hyponasality</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>Cul-de-sac resonance</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>Audible nasal emission</td>
<td>frequent</td>
<td>absent</td>
<td>occasional</td>
<td>occasional</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>frequent</td>
<td>occasional</td>
<td>absent</td>
</tr>
<tr>
<td>Nasal turbulence</td>
<td>absent</td>
<td>absent</td>
<td>frequent</td>
<td>frequent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>Voice disorder</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
<td>absent</td>
</tr>
<tr>
<td>Speech intelligibility</td>
<td>mild</td>
<td>normal</td>
<td>moderate</td>
<td>moderate</td>
<td>severe</td>
<td>severe</td>
<td>severe</td>
<td>severe</td>
<td>mild</td>
<td>normal</td>
</tr>
</tbody>
</table>

| Objective resonance measurements    |               |               |               |               |               |               |               |               |               |
| Nasalance (%)                       | Norm*         |               |               |               |               |               |               |               |               |
| Vowel /u/                           | 17±11.1       | 2             | 39            | 47            | 36            | 52            | 75            | 38            | 45            | 14            | 18            |
| Bilabial sentences                  | 162.7±41      | 6             | 69            | 53            | 67            | 62            | 39            | 51            | 10            | 8             |
| Alveolar sentences                  | 152.7±39      | 7             | 70            | 67            | 66            | 64            | 39            | 51            | 16            | 13            |
| Velar sentences                     | 19±8.7        | 8             | 69            | 52            | 69            | 57            | 41            | 55            | 13            | 12            |
| Sibilant sentences                  | 18±9.0        | 7             | 78            | 68            | 76            | 72            | 42            | 70            | 79            | 20            |
| Nasal sentences                     | 64±11.3       | 7             | 74            | 65            | 81            | 77            | 50            | 66            | 80            | 68            |
| Rainbow passage                     | 33±7.3        | 28            | 72            | 72            | 71            | 65            | 39            | 59            | 66            | 34            |
| Zoo passage                         | 14±6.4        | 13            | 70            | 65            | 66            | 62            | 37            | 55            | 49            | 14            |
| VLHR /i/ 4.47*F0Hz (dB)             | 25.48±10.30   | 23.37±24.41   | 25.29±23.16   | 23.62±24.50   | 20.55±20.15   |               |               |               |               |               |
| NSI 2.0                             | -3.5±4.1      | -14.5±12.5    | -14.3±14.7    | -5.2±10.7     | -5.6±3.2      |               |               |               |               |               |

| Articulation                        | W | S | W | S | W | S | W | S | W | S | W | S | W | S | W | S | W | S | W | S |
| % consonants correct                | 46| 48| 67| 68| 66| 39| 66| 40| 47| 38| 71| 49| 48| 26| 36| 33| 70| 59| 84| 73|
| % plosives correct                  | 41| 46| 73| 78| 72| 30| 57| 28| 28| 27| 63| 57| 40| 9 | 19| 16| 75| 72| 81| 67|
| % fricatives correct                | 15| 4 | 49| 17| 26| 0 | 46| 4 | 12| 4 | 46| 0 | 3 | 4 | 0 | 0 | 24| 4 | 75| 70|
| % /t,d/ correct                     | 5 | 12| 33| 41| 62| 24| 57| 24| 14| 24| 43| 59| 38| 6 | 19| 38| 38| 29| 33| 12|
| % /s,z/ correct                     | 3 | 0 | 48| 14| 3 | 0 | 30| 4 | 0 | 0 | 39| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 67| 68|

W = Word level, S = Sentence level; *Luyten et al., 2012
Therefore, this patient was referred to discuss the possibility of speech improvement surgery, such as buccal (Mann et al., 2011) or pharyngeal flap surgery (Sloan, 2000). Only when an appropriate structure and functioning of the velopharyngeal mechanism is achieved, amelioration of the resonance and articulation, supported by speech therapy to reduce persisting errors, may be expected.

Regarding the other four patients, two patients (2 and 3) still showed speech disorders that required additional therapy sessions. Two patients (1 and 5) presented with minimal articulation errors and negligible resonance disorders resulting in normal speech intelligibility, thus negating the requirement for further treatment. Potentially, the speech intelligibility improvements of patients 1 and 5 might be explained by a combination of a young age at palatal closure (≤ 12 months), a young age at speech therapy (< 12 years), very good proficiency of the language used during treatment and a limited number of affected consonants. Considering this last factor, compensatory articulation was only observed in the consonants /s,z/ in these patients. Consequently, therapy only had to focus on these sounds. Additionally, patient 1 rapidly achieved a correct direction of the oral airflow during the production of all consonants only by modeling the buildup of more intra-oral pressure during articulation. The fact that she already followed 12 hours of speech therapy may also have contributed to this fast acquisition. In patients 2 and 3, more consonants were affected and their speech was perceived to be less intelligible. Therefore, only a selection of target consonants could be treated during this limited amount of therapy sessions, resulting in the need for more speech therapy. However, some transmission of the learned principles on consonants that were not treated could be observed as nasalance values for stimuli including those consonants also decreased. A possible explanation for this generalization is that these patients were able to transfer the skill of redirecting airflow learned with the target consonants to untreated consonants.

One of the main limitations of this study was that the patients’ closure pattern of the velopharyngeal sphincter could not be verified by nasopharyngoscopy and/or videofluoroscopy prior to speech therapy, as the necessary equipment was not available in CoRSU. Particularly for patient 4, velopharyngeal insufficiency was presumed to explain the limited progress during speech therapy. Another limitation of this study includes the language in which the speech therapy was provided. As English was the second language in four of the five patients, therapy could possibly have been more effective when it was provided in the patients’ mother tongue. Moreover, transfer of the improved sound productions to the mother language could not reliably be verified as no SLP was available who spoke the mother language of the patients. Additionally, given that all patients spoke another mother language and nasalance scores can be influenced by dialect (Awan et al., 2015; Mayo et al., 1996; Nichols, 1999; Seaver et al., 1991), the comparison of the obtained nasalance with the available normative values for Ugandan-English speaking children (Luyten et al., 2012) have to be done with care. However, overall perceptual judgments were comparable with the instrumental measurements.
Due to organizational reasons, only six hours of therapy could be provided to a limited number of patients, which is a reduction compared to previous ‘speech camp’ studies (Pamplona et al., 2005 – 60h; Prathanee et al., 2011 – 18h + 6h follow-up; Van Demark & Hardin, 1986 – 104h) and prevented statistical testing to confirm significant improvements after therapy. An additional limitation is the lack of follow-up data at this moment. Although preliminary promising results were achieved in the current study, three patients needed further speech therapy, whether or not in combination with speech improving surgery. As our research unit can only spent short periods of time in Uganda, additional solutions were devised to ensure sustainability. First, speech exercises were offered to all treated patients to practice at home. Additionally, the relatives of the patients were invited to follow the therapy sessions so that information could be given about the basic therapy materials in combination with instructions for follow-up (D’Antonio & Nagarajan, 2003). However, only two mothers (patient 1 and 5) attended the sessions and this solution was only applicable on the patients included in this study. Nevertheless, the results of this pilot study encouraged the local management to explore the opportunities to set up a speech-language therapy unit at CoRSU. A possibility to expand the access to speech-language therapy services at CoRSU is the training of other, available professionals (e.g. nurses, physiotherapists, etc.) in the principles of speech-language pathology (D’Antonio & Nagarajan, 2003) pending the availability of local Ugandan SLPs on the long term. Meanwhile, follow-up of the previously treated patients is necessary and will be provided during subsequent missions. Whether the established progress in speech would remain after several months is subject for further research.

Additionally, it should be assessed whether the current treatment protocol can replace traditional therapy schedules in countries with sufficient access to speech-language therapy. The effect of intensive treatment is already explored in other speech therapy domains, such as voice (Fu et al., 2014; Wenke et al., 2014), stuttering (Euler et al., 2014), aphasia (Cherney et al., 2008; Harnish et al., 2014), speech and voice in persons with Parkinson’s disease (Spielman et al., 2007) and speech sound disorders not due to cleft palate (Allen, 2013), in which comparable or superior results were found for intensive therapy schedules compared with a traditional therapy frequency. Promising results are also reported by Pamplona et al. (2005), who revealed that short and intensive speech therapy during a speech summer camp of three weeks (± 60 therapy sessions of one hour) resulted in similar articulation outcomes compared to twelve months of traditional therapy (± 104 therapy sessions of one hour) in children born with C(L)P. Moreover, cost-effectiveness of a four-day speech camp and follow-up session was proven by Prathanee et al. (2011). Considering that an increase in therapy frequency can induce a reduction of the requested amount of therapy sessions, health insurance expenses, financial contributions of the parents and patients’ tiredness of therapy can decrease. These advantages are applicable to speech therapy services in resource-poor as well as in developed countries, which encourages further research to answer this question reliably.
Conclusion

In conclusion, short and intensive speech therapy (six hours during three or four days) provided to carefully selected Ugandan patients with C(L)P resulted in an improvement of articulation and resonance, although additional therapy sessions were often required (3/5, 60%). Consequently, this treatment approach is considered to be effective and feasible in countries with limited access to speech-language therapy, when follow-up therapy sessions are ensured. Future research should allow for the above-mentioned limitations, including direct visualization of the velopharyngeal mechanism prior to speech therapy and the availability of Ugandan professionals speaking local languages who are trained in the principles of speech-language therapy. Whether this treatment protocol should replace the traditional therapy models in countries with sufficient access to speech-language therapy is subject for further research.

Acknowledgements

The authors gratefully acknowledge all participants for their contribution as well as Sofie Degeest and Paul Corthals for their support in the audiological assessments.
References


PART 3

GENERAL DISCUSSION
CHAPTER 9

GENERAL DISCUSSION AND CONCLUSION

For decades, researchers have been searching for the most ideal assessment technique in order to diagnose resonance disorders and to decide on the most appropriate treatment. Determining the presence and degree of resonance disorders is essential during diagnosis or treatment evaluation as this influences further therapy policies. The introduction of this doctoral thesis provided a profound description and analysis of the advantages and limitations of the current assessment techniques. On the one hand, accurate information based on direct techniques is needed to diagnose the amount of velopharyngeal dysfunction. On the other hand, resonance disorders should be determined by using perceptual assessment and indirect techniques. However, several variables can influence listeners’ perception of speech which may limit the reliability and validity of perceptual judgments (Kent, 1996; Kreiman et al. 1993). Consequently, numerous indirect assessment techniques were developed to complement and objectify perceptual assessments. Nevertheless, no indirect technique can yet closely reflect the capabilities of the human ear.

A possible solution to sidestep the limitations of single indirect instrumental assessment techniques is the combination of different variables into a multiparametric index that allows to implement complementary indirect measurements. Considering this, the initial aim of this doctoral thesis was to explore the application of the Nasality Severity Index (Van Lierde et al., 2007), as a new, multiparametric approach to determine hypernasality in daily clinical practice.

To verify the possible influence of personal and environmental variables on the original NSI, the data of a larger control group than the one included in the pilot study performed by Van Lierde et al. (2007) were analyzed in chapter 3. Additionally, the availability of reference values for Dutch-speaking Flemish children without resonance disorders for this index was aimed. Next to the influence of age due to the inclusion of the maximum duration time (MDT) of /s/, large standard deviations of the mean were observed for the NSI. This resulted in large 95% prediction intervals (95% PI boys [-19.3;+35.5]; girls [-30.6;+43.7]), indicating a great spread of NSI values in children without resonance disorders, even in children at the same age. When comparing these 95% PIs with the 95% PIs of the NSI scores obtained from the patients with perceived hypernasality included in the pilot study by Van Lierde et al. (2007) (mean=-24.0, SD=21.9, 95% PI [-67.9;+19.9]), a large overlap was observed. Due to these large differences, the original NSI could no longer be considered a reliable instrument to assess the resonance of an individual person. Moreover, concerns were introduced about the inclusion of MDT of /s/ and the mirror-fogging test by Glatzel regarding the possible influence of personal and environmental factors. Therefore, an adaptation of the original NSI
with inclusion of new assessment techniques was recommended in which the influence of personal and environmental variables had to be reduced.

Based on the rationale of a multiparametric index as an instrumental correlate for hypernasality and the recommendations of the above-mentioned study, additional acoustic techniques to determine hypernasality were explored in chapter 4. In addition to the nasalance values of different stimuli obtained with a Nasometer (model 6450, KayPentax), one-third octave spectrum analysis (Kataoka et al., 1996) and voice low tone to high tone ratio (VLHR) (Lee et al., 2003) were described as spectral analyses based on the acoustic features of hypernasality. Based on the optimal statistical discrimination of 35 children with perceived hypernasality and a control group of 50 children without resonance disorders, a weighted linear combination of three acoustic parameters was established, using a stepwise statistical approach (i.e. logistic regression analysis) with sensitivity and specificity as the serving criteria. More specifically, the nasalance value of the vowel /u/ and an oral text passage obtained by a Nasometer, and VLHR of a sustained vowel /i/ with a cutoff frequency of 4.47*F0Hz (originally described by Lee et al. (2003)) were included. The formula of the adapted NSI, the Nasality Severity Index version 2.0, yields NSI 2.0 = 13.20 – (0.0824 x nasalance /u/ (%)) – (0.260 x nasalance oral text (%)) – (0.242 x VLHR /i/ 4.47*F0Hz (dB)). With a sensitivity of 92% and a specificity of 100%, using a cutoff score of zero, the NSI 2.0 distinctively discriminates children with hypernasality from children with normal resonance, in which patients with perceived hypernasality have scores below zero. Moreover, the index identifies patients and control children more accurately in comparison with the results from the nasalance score for an oral text in itself (81% sensitivity, 93% specificity), which proves the preference of a multiparametric index over a single parameter approach. Ultimately, the validity of the NSI 2.0 was proven to be high when applying the derived formula on the data of an independent Ugandan patient and control group speaking English as a second language (sensitivity 88%, specificity 89%).

One of the necessary conditions to implement this new index in daily clinical practice and to evaluate interventions properly, is the availability of normative values derived from children and adults without resonance disorders. To formulate these reference values, the possible influence of gender and age on the NSI 2.0 was explored in chapter 5, to verify the need for separate reference values according to gender and/or age. Based on the data of 80 normal-developing Dutch-speaking Flemish children between 4 and 12 years old, no influence of gender or age on the NSI 2.0 scores was found. However, based on the data of 60 normal-Dutch-speaking Flemish adults between 18 and 60 years old, significantly higher NSI 2.0 scores were observed in men compared to women, without any age effect. When comparing the data of children and adults, a significant interaction between gender and age was found, in which adult men showed higher NSI 2.0 scores compared to adult women and children. Based on these study outcomes, separate reference values for the NSI 2.0 and its parameters were established for children, adult men and adult women. All participants’ scores exceeded zero (range +0.1 to +7.3 in children, range +0.6 to +8.7 in adults), which is in
accordance with the cutoff score of zero described in chapter 5. Moreover, a smaller standard deviation of the children’s scores (SD 1.63) was observed compared with the standard deviations of the original NSI (SD 13.8 in boys and SD 18.8 in girls, chapter 3).

As the stability of instrumental measurements can be affected by several sources such as instrumental variance, test procedure, subject performance and environment, the reliability of the NSI 2.0 was verified in chapter 6. More specifically, both the short-term and long-term test-retest reliability of the NSI 2.0 and its components was explored in 29 children and 40 adults without resonance disorders. A comprehensive set of statistical measures, including the intra-class correlation coefficient (ICC), the standard error of measurement (SEM) and the minimal detectable difference (MDD), revealed a somewhat larger long-term variability of the NSI 2.0 and its parameters compared to the short-term variability, both in children and adults. Overall, for children, a difference of 2.68 between the results of two consecutive measurements can be interpreted as a genuine change. For adults, the difference is 2.82. The application of these findings was illustrated by presenting two clinical case studies. With an ICC of 0.77 in children and 0.84 in adults, the NSI 2.0 additionally shows an excellent relative consistency. Differences between two measurements may be explained by a random error caused by repositioning the Nasometer headgear (Lewis et al., 2008; Watterson et al., 2005) and intra-subject variability due to the variation in performance and physiological factors such as small changes in nasal patency (de Boer & Bressmann, 2014; Lewis et al., 2008; van Doorn & Purcell, 1998). The smallest agreement was found for the parameter ‘VLHR of /i/’. Although the influence of noise and loudness variations was expected to be limited by using a relative rather than an absolute index (Lee et al., 2003), this higher variability may indicate the importance of controlling for these influencing factors.

With a high sensitivity and specificity, the NSI 2.0 accurately identifies hypernasality in patients with a history of cleft palate (chapter 4). However, the final purpose of the NSI 2.0 was to provide an easy-to-interpret severity score of hypernasality to facilitate the evaluation of therapy outcomes, communication to the patient and other clinicians, and decisions for treatment planning. Therefore, the correlation between the NSI 2.0 scores of patients with hypernasality and the perceptual judgment of hypernasality based on spontaneous speech and sentence repetition was explored in chapter 7. Additionally, the possible influence of audible nasal airflow and speech intelligibility on the NSI 2.0 scores was investigated. Based on the results of this study, the NSI 2.0 scores correlate significantly with the perceptual ratings of hypernasality (r=-0.64), in which a more negative NSI 2.0 score indicates the presence of more severe hypernasality. However, a comparable correlation was found for the nasalance scores of an oral text alone. A possible explanation may be that the additional parameters included in the NSI 2.0 (i.e. nasalance score of /u/ and VLHR of /i/) are only based on single vowels which may explain their low correlations with the perception of hypernasality in continuing speech. Furthermore, a multiple regression analysis withheld only the amount of perceived hypernasality to explain the variance in NSI 2.0 scores.
Therefore, no significant additional influence can be expected on the index from the variables **audible nasal emission** and **speech intelligibility**. As the NSI 2.0 correlates significantly with perceived hypernasality, it provides an easy-to-interpret severity score of hypernasality which will facilitate the evaluation of therapy outcomes, communication to the patient and other clinicians, and decisions for treatment planning.

Finally, the **NSI 2.0 was applied to objectify the short-term effectiveness of short, intensive speech therapy on the resonance of patients with a history of cleft (lip and) palate (C(L)P)** in chapter 8. Therefore, five Ugandan patients (age: 7;4-19;7 years) with non-syndromic C(L)P received six hours of individualized speech therapy in three or four days in which the therapy focused on correct phonetic placement and contrasts between oral and nasal airflow. The NSI 2.0 scores of the two patients who were initially judged with hypernasality or audible nasal emission and post-therapy with normal resonance changed from a negative score pre-therapy to a positive score post-therapy. A change of more than 2.68 points on the NSI 2.0 scores was observed, which reflects this genuine change (chapter 6). Two other patients were judged with severe hypernasality before and after therapy. As a difference of less than 2.68 points was observed in the NSI 2.0 score of one of these patients, this is in accordance with the identical perceptual judgments pre- and post-therapy. However, a decrease in the NSI 2.0 score of more than 2.68 points was noticed in the other patient. As this patient was already judged with the maximum grade of hypernasality, this grade could not be increased anymore. Still, the frequency of audible nasal emission increased from absent pre-therapy to frequently observed post-therapy. Although the presence of audible nasal airflow was not withheld as an influencing variable on the NSI 2.0 in chapter 7, the increase of audible nasal emission may have contributed to the decrease in the NSI 2.0 score for this patient. The fifth patient presented with severe hypernasality pre-therapy and moderate hypernasality post-therapy. However, the NSI 2.0 scores did not reflect this perceptual change. Based on this pilot study, the contribution of the NSI 2.0 to evaluate therapy effects is not straightforward. However, as only five patients were included, the representativeness of these findings may be limited which supports the need for further research.

Due to the focus on the NSI 2.0 to evaluate hypernasality before and after speech therapy, the impression may be created that speech therapy alone can always eliminate hypernasality. However, speech therapy in this study focused on both, eliminating compensatory articulation errors and redirecting oral and nasal airflow which had an indirect positive effect on resonance. Yet, in case of velopharyngeal insufficiency, surgery is required to repair the anatomical defect after which speech therapy can be applied to remedy persisting articulation and airflow errors. The study described in chapter 8 originally aimed to evaluate the short-term effectiveness of short and intensive speech therapy provided to patients with C(L)P in terms of both, articulation and resonance. Based on perceptual and instrumental assessments before and after therapy in five Ugandan patients, speech improvements were observed for most patients, although to varying degrees. Clinically
relevant progress of objective nasalance values and/or articulation was obtained in four patients, resulting in two patients who showed normal speech intelligibility after three or four days of speech therapy, and three patients who required additional therapy. These preliminary short-term results demonstrated that short and intensive speech therapy can be effective for patients with C(L)P in countries with limited access to speech-language therapy.

**SWOT analysis of the NSI 2.0**

Table 9.1 presents a SWOT analysis of the NSI 2.0, including its strengths, weaknesses, opportunities and threats. Based on this analysis, future perspectives will be discussed.

**Strengths**

Reducing the results of different indirect assessment techniques into one NSI 2.0 value sidesteps the limitations of a single parameter approach, but preserves the distinctness of the interpretation of a single value, which is especially useful in the communication with patients, their relatives and clinicians other than speech-language pathologists. Due to the inclusion of different parameters, the NSI 2.0 discriminates children with perceived hypernasality from children without resonance disorders with a high sensitivity and specificity. Additionally, a significant correlation was found between the NSI 2.0 scores and the perceptual ratings of an expert panel in which a more negative NSI score indicates the presence of more severe hypernasality. As reliable test-retest measurements were observed, this new multiparametric index can be used in the follow-up of patients with perceived hypernasality, e.g. to verify the effect of surgery in case of velopharyngeal insufficiency. Finally, the results of a patient can be compared to the available reference values obtained from Dutch-speaking Flemish children and adults without resonance disorders to verify the (remaining) amount of deviation from normal resonance during diagnosis. Based on this information, decisions about the need for intervention or termination of treatment, and follow-up can be made.
Table 9.1 Strengths, weaknesses, opportunities and threats (SWOT) analysis of the Nasality Severity Index 2.0.

<table>
<thead>
<tr>
<th>STRENGTHS</th>
<th>WEAKNESSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>-Noninvasive, convenient</td>
<td>-No information about additional resonance disorders</td>
</tr>
<tr>
<td>-Easy to interpret</td>
<td>-No information about speech intelligibility and acceptability</td>
</tr>
<tr>
<td>-One measurement value results in clear communication</td>
<td>-Possible influence of audible nasal emission and turbulence on NSI 2.0 score</td>
</tr>
<tr>
<td>-High sensitivity and specificity</td>
<td>-Only 39% of variance in NSI 2.0 scores can be explained by the amount of hypernasality</td>
</tr>
<tr>
<td>-Significant correlation with perceptual judgments</td>
<td></td>
</tr>
<tr>
<td>-Good test-retest reliability</td>
<td></td>
</tr>
<tr>
<td>-Reliable measurements offer the possibility for reliable follow-up</td>
<td></td>
</tr>
<tr>
<td>-Reference values available for Dutch-speaking Flemish children and adults</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>OPPORTUNITIES</td>
<td>THREATS</td>
</tr>
<tr>
<td>-Higher sensitivity and specificity compared to single parameter approach</td>
<td>-Comparable correlations with perceptual judgments for the NSI 2.0 scores and nasalance scores based on oral text alone</td>
</tr>
<tr>
<td>-Quantitative results → bias due to listener characteristics is reduced</td>
<td>-Resistance to use computer-based analyses</td>
</tr>
<tr>
<td>-Application in other patient populations possible</td>
<td>-Need for Nasometer → expensive, not always available</td>
</tr>
<tr>
<td>-Further exploration and potential inclusion of computer-based acoustic analyses possible</td>
<td>-Development of acoustic analyses by other research groups (de Boer &amp; Bressmann, 2015a, 2015b)</td>
</tr>
<tr>
<td></td>
<td>-Perceptual judgments remain gold standard</td>
</tr>
</tbody>
</table>

Weaknesses

Since the NSI 2.0 was developed as an indirect measure of hypernasality, no nasal speech stimuli are included in the index which prevents the index to detect hyponasality or mixed nasality. Although hypernasality affects speech intelligibility and acceptability more than hyponasality does and therefore is clinically more relevant (Shprintzen et al., 1979), some authors highlight the need to identify the amount of hyponasality and mixed nasality. This results from the evidence that hypernasal resonance in combination with reduced nasal patency may occur among patients with cleft lip and palate due to a septum deviation, narrow vestibule, maxillary retraction or after speech improving surgery such as a pharyngeal flap (Fukushiro & Trindade, 2011; Kummer, 2011; Nellis et al., 1992; Shprintzen et al., 1979) or among patients with apraxia of speech who have difficulties with opening and closing the velopharyngeal valve at an appropriate time (Kummer, 2011). Additionally, no information about the intelligibility and acceptability of a patient’s speech can be derived from the NSI 2.0.

Furthermore, the influence of audible nasal emission and turbulence on the NSI 2.0 scores is not straightforward. Although the amount of audible nasal airflow was not withheld as an additional influencing variable to explain the variance in NSI 2.0 scores by a multiple
regression analysis (chapter 7), the increase of audible nasal emission in a patient included in
the study presented in chapter 8 may have contributed to the decrease in the NSI 2.0 score
for this patient. Moreover, a second patient who presented with audible nasal emission
without hypernasality before therapy had a negative NSI 2.0 score, incorrectly indicating the
presence of hypernasality. After therapy, her score exceeded zero, indicating no
hypernasality. However, the increase was due to the absence of audible nasal emission after
therapy. Although the possible influence of audible nasal airflow on the NSI 2.0 may be due
to the inclusion of the nasalance score of an oral text as a parameter, these observations are
only based on two patients which may limit their representativeness and support the need
for further research. Nevertheless, as a Nasometer cannot discriminate between acoustic
energy from nasal resonance and energy from aerodynamic phenomena (such as audible
nasal emission and turbulence), nasalance scores may be increased in patients with audible
nasal airflow problems (Karnell, 1995; Sweeney & Sell, 2008; Watterson et al., 1998). The
inclusion of only low-pressure consonants may limit the influence of audible nasal airflow as
this phenomenon particularly occurs in high-pressure consonants (Watterson et al., 1998).
However, the exclusion of high-pressure consonants may result in a stimulus that is less
representative for spontaneous speech. In conclusion, all the above-mentioned limitations
highlight the persistent need to complement instrumental measures with perceptual
assessments.

Another limitation of the NSI 2.0 is that only 39% of the variance in NSI 2.0 scores could be
explained by the amount of perceived hypernasality. This may be due to restrictions of the
included acoustic analyses to represent the perceived hypernasality and/or restrictions of
the perceptual assessments due to task factors (e.g. applied rating scale and perceptual
context) and individual differences of the listeners (e.g. experience, attention laps) as
discussed in chapter 7.

**Opportunities**

One of the opportunities of the NSI 2.0 compared to a single parameter approach,
such as the nasalance score of an oral text alone, is its higher sensitivity and specificity to
identify patients with hypernasality and subjects without resonance disorders (chapter 4 and
7).

Furthermore, compared to perceptual judgments, NSI 2.0 scores are based on quantitative
information which reduces the possible bias of listener characteristics. As reported by Kent
(1996), several studies caution for the influence of a patient’s physical features and the
knowledge of the patient’s history on perceptual judgments. Especially during the evaluation
of intervention, this self-fulfilling prophecy phenomenon can induce bias. Additionally, the
listener’s familiarity with a specific speaker can influence the perception as there is evidence
that listeners can adapt to atypical patterns of speech production (Kent, 1996). As a result,
NSI 2.0 scores can be helpful to make reliable comparisons between different assessment
moments.
Currently, NSI 2.0 scores were analyzed for patients with hypernasality due to different causes, i.e. cleft (lip and) palate, submucous cleft palate, after adeno(tonsillec)otomy, in patients with velo-cardio-facial syndrome and due to velopharyngeal mislearning (chapter 7). The opportunity exists to apply the NSI 2.0 in patients with hypernasality due to, for example, velopharyngeal incompetence (e.g. dysarthria and apraxia of speech) or velopharyngeal mislearning in case of hearing impairment or deafness.

Finally, thanks to the current trend of digitalization, the exploration and development of computer-based acoustic analyses to detect and measure resonance disorders is continuously expanding. Due to the multiparametric character of the NSI, adaptations to new techniques can easily be made by adding or replacing (some of) the parameters. As a result, the NSI can evolve together with new insights in this domain and become an even more reliable and valid indirect measure of hypernasality including the most recent measurement techniques.

**Threats**

One of the threats which may prevent the application of the NSI 2.0 in daily clinical practice is the comparable correlation with perceptual judgments of hypernasality found for the NSI 2.0 scores and the nasalance values of an oral text alone. Although it was hypothesized that a multiparametric index should correlate better with the perception of hypernasality, the inclusion of parameters based on isolated vowels only in addition to the nasalance value of an oral text may have prevented to confirm this hypothesis (chapter 7). Nevertheless, the inclusion of these parameters into the NSI 2.0 contributes to the better identification of patients with hypernasality in comparison with a single parameter approach based on the nasalance scores of an oral text alone (chapter 4 and 7).

A second threat for the implementation of the NSI 2.0 is a possible resistance among speech-language pathologists to use computer-based analyses as this is often perceived as time-consuming or difficult. Therefore, a script was created to determine the VLHR score of /i/ and the NSI 2.0 score (see appendix chapter 7), in analogy to the script to determine the Acoustic Voice Quality Index (AVQI) (Maryn, 2013; Maryn et al., 2010). After recording the vowel /i/ using PRAAT software, the examiner only needs to select a 0.5s sample of the recording and run the script. As the script asks to fill in the nasalance scores of /u/ and an oral text, the NSI 2.0 score is immediately provided together with the VLHR score. Since the analysis can be performed in real time and is not time-consuming, the results can be discussed immediately with the patient.

A third threat is the need for specialist equipment, more specifically a Nasometer, to determine the NSI 2.0. A recent survey in North America (Stelck et al., 2011), including answers of 38 speech-language pathologists specialized in the treatment of resonance disorders, revealed that 42.1% had never used a Nasometer of whom 68.8% did not have access to the instrument. Although a Nasometer is often available in tertiary-care hospitals,
the instrument remains very expensive and is often not available for speech-language pathologists.

An additional threat for the NSI 2.0 is the development of similar acoustic analyses based on a combination of different parameters. de Boer and Bressmann (2015b), for example, developed two formulas based on nasalance values of an oral and nasal stimulus that predict a resonance condition (i.e. normal resonance, hypernasality, hyponasality, mixed nasality) with an overall efficacy of 88.6% based on simulations of those four conditions made by 11 female participants without resonance disorders. Additionally, they analyzed the long-term averaged spectra (LTAS) of the same oral and nasal speech stimuli to derive two formulas using only acoustic analyses based on PRAAT software, resulting in the correct classification of 80.7% of the simulations (de Boer & Bressmann, 2015a). Although these preliminary results are promising, further research, including patient data, is necessary to verify the possible influence of articulation errors and audible nasal airflow on the formulas. Furthermore, only simulations of severe hypernasality were included, which does not reflect the complete range of hypernasality in patients. Nevertheless, the reasoning to include an additional index based on nasal stimuli to create a two-dimensional approach (hypernasality and hyponasality) of oral-nasal balance is interesting.

Finally, as speech perception is perceptual in nature and the NSI 2.0 still does not completely represent the perceptions of the human ear, the index may be perceived as inferior to perceptual judgments (Moll, 1964). However, perceptual judgments are also still vulnerable to influencing factors which limits their reliability and validity (Kent, 1996; Kreiman et al., 1993). As both assessment procedures are complementary, they can restrain each other’s limitations and may stimulate critical thinking, especially when contradictory results are observed.

**Future perspectives**

Based on the SWOT analysis, future perspectives can be formulated to amplify the strengths and opportunities of the NSI 2.0 and to remedy its weaknesses and threats.

Based on the current study, reference values for the NSI 2.0 are now available for Dutch-speaking Flemish children and adults (chapter 5). However, further research is required to explore the need for reference values in different languages, given the influence of language on nasalance scores (Nichols, 1999; Okalidou et al., 2011; Rochet et al., 1998; Van Lierde et al., 2001).

So far, only data of participants without resonance disorders were analyzed to determine the reliability of the NSI 2.0 (chapter 6). However, test-retest variability may be larger in patients compared to normal-speaking participants. In their study, Watterson and Lewis (2006) found nasalance score differences up to 10% for 94% of the consecutive measurements of an oral text in a short-term condition with replacement of the headgear including patients with cleft palate (mean age 10y8m, 3y3m-26y). These differences are
higher compared to the nasalance score differences up to 5% for 92% (Lewis et al., 2008) and 95% (Watterson et al., 2005) of the consecutive measurements of an oral text reported for participants without resonance disorders. Therefore, further research could focus on the reliability of the NSI 2.0 in patients with hypernasality.

One of the limitations of the NSI 2.0 is that comparable correlations with perceptual judgments were found for the NSI 2.0 and the nasalance score of an oral text alone. Additionally, only limited correlations were found between the other parameters of the NSI 2.0 and the amount of perceived hypernasality. As discussed in chapter 7 and the SWOT analysis above, these two additional parameters are only based on isolated vowels which may have caused these limitations. Recently, de Boer and Bressmann (2015a) introduced the analysis of LTAS on continuous speech to identify simulation-based disorders of oral-nasal balance. LTAS provides information about the distribution of energy throughout the speech spectrum. As the presence of hypernasality is characterized by a redistribution of energy in the speech spectrum due to the coupling of the nasal tract (Chen, 1997; Fant, 1970; Hawkins & Stevens, 1985; Schwartz, 1968), LTAS was applied to detect hypernasality in connected speech. Using bandwidths of 100Hz within a frequency spectrum of 0 to 4000Hz, six frequency bands of an oral speech stimulus and four frequency bands of a nasal speech stimulus were selected to discriminate between four speech conditions (i.e. normal resonance, hypernasality, hyponasality, mixed nasality). With an 80.7% correct classification of the simulated speech samples, these first results are promising. Therefore, further research is already initiated by our research group to explore the application of LTAS analysis on connected speech in patients with perceived hypernasality and a control group without resonance disorders to verify its discriminatory value and its relation with perceived degree of hypernasality. The application of the VLHR approach in addition to the LTAS analysis could be explored too, as the summation of energy over the full frequency spectrum between 0 and 4000Hz instead of a selection of frequency bands may hypothetically result in an even better classification. Another possibility to improve the representativeness of spectral analyses based on vowels is the extraction and concatenation of the vowels /i/ from different words or sentences. Based on this technique, the analysis can be applied on a more extensive sample and the influence of coarticulation can be taken into account. However, as extracting vowels /i/ from words or sentences may be time-consuming, this technique may be less useful in clinical practice. Furthermore, the need for an additional index including a nasal speech stimulus to meet the idea of a two-dimensional approach of oral-nasal balance disorders, as suggested by de Boer and Bressmann (2015a, 2015b), could be investigated too. As the parameters of the NSI 2.0 can be easily adapted, the NSI can evolve together with the exploration and development of these new techniques. As a supplementary result, an index including only parameters based on acoustic analyses using free available software such as PRAAT (Boersma & Weenink, 2014) may be established, which may eliminate the need for specialist equipment, more specifically a Nasometer.
Finally, several authors consider perceptual assessments as the standard against which instrumental measurements must be validated (Baylis et al., 2015; Kent, 1996; Keuning et al., 2004; Moll, 1964; Vogel et al., 2009). However, since perceptual assessments are often influenced by errors and bias, the reliability and validity of this validation can be questioned (Kent, 1996). Recently, researchers started focusing on the development of standard protocols to assess speech and intervention outcomes, including standard speech samples, recording methods, speech analyses and training of listeners (Chapman et al., 2016; Henningsson et al., 2008; John et al., 2006; Lohmander et al., 2009) as described in chapter 1. Although encouraging results are reported regarding the intra- and inter-rater reliability and validity, questions remain about, for example, the applied rating scale (Baylis et al., 2015; Whitehill et al., 2002). Therefore, further research also needs to focus on the optimization of perceptual assessment protocols in addition to the improvement of instrumental measurements to result in the most reliable and valid approach of resonance disorders in clinical practice and research. Consequently, both instrumental and perceptual assessments can be applied in the evaluation of the effectiveness of an intervention which contributes to the persistent need for evidence based practice in the domain of resonance.

In this view, the short-term effect of intensive speech therapy on the articulation and resonance in five Ugandan patients with cleft (lip and) palate was explored in chapter 8. Based on this pilot study, the influence of audible nasal airflow on the NSI 2.0 scores and the contribution of the NSI 2.0 to evaluate therapy effects on hypernasality was not straightforward. However, as these conclusions are only based on the data of five patients, their representativeness can be questioned. Therefore, further research could focus on collecting data of a more extensive patient group, resulting in the possibility to perform a statistical analysis to draw more reliable conclusions. Additionally, perceptual and instrumental data collection before and after other therapy approaches, such as speech improving surgery, can be performed. Based on perceptual judgments by experienced SLPs who are blind for the study purpose, the sensitivity of the NSI 2.0 to detect possible changes in hypernasality due to intervention can be evaluated.

Furthermore, future research can investigate the long-term effect of the intensive therapy approach proposed in chapter 8. Additionally, it could be assessed whether the applied treatment protocol can replace traditional therapy schedules in countries with sufficient access to speech-language therapy. Considering that an increase in therapy frequency can induce a reduction of the requested amount of therapy sessions, health insurance expenses, financial contributions of the parents and patients’ tiredness of therapy can decrease. These advantages are applicable to speech therapy services in resource-poor as well as in developed countries, which encourages further research to answer this question reliably.
Conclusion

Considering that no indirect assessment technique was yet available that highly correlates with the perception of hypernasality as detected by the human ear, the initial aim of this doctoral thesis was to explore the application of the Nasality Severity Index, originally developed by Van Lierde et al. (2007), as a new, multiparametric approach to determine hypernasality in daily clinical practice. Based on the studies described in this work, the Nasality Severity Index 2.0 was created. In order to enable its application in clinical practice, the following research objectives were achieved:

- The possible influence of personal and environmental variables on the original NSI, developed by Van Lierde et al. (2007), was explored, resulting in the adaptation of the original NSI into the NSI 2.0.
- The NSI 2.0 discriminates children with perceived hypernasality and control children without resonance disorders with a higher sensitivity and specificity in comparison with a single parameter approach.
- Reference values for Dutch-speaking Flemish children and adults are now available for the NSI 2.0.
- Short-term and long-term reliability of the NSI 2.0 was verified in normal-Dutch-speaking Flemish children and adults, in which a difference of respectively 2.68 and 2.82 in NSI 2.0 scores obtained from two consecutive measurements can be interpreted as a genuine change.
- The correlation between NSI 2.0 scores of patients with hypernasality and perceptual assessment of hypernasality based on spontaneous speech was explored resulting in a correlation of -0.64. No additional influence of audible nasal airflow or speech intelligibility on the variance of NSI 2.0 scores was withheld.
- The NSI 2.0 can be applied to verify the effect of intervention on resonance in patients with a history of cleft palate in addition to perceptual evaluations.

In conclusion, the NSI 2.0 is a quantitative identifier of hypernasality based on a multiparametric approach using two different acoustic measurement techniques. This results in an index with a higher sensitivity and specificity compared to a single parameter approach, that is noninvasive, easily repeatable, and convenient to determine and interpret. Furthermore, limitations of the original NSI procedure (Van Lierde et al., 2007), such as influence of environmental and personal variables, were restrained and a good test-retest reliability was found. Future research can explore additional instrumental correlates of perceived hypernasality based on connected speech, e.g. LTAS analysis or VLHR, which may eventually result in the elimination of the need for a Nasometer. In conclusion, the multiparametric NSI 2.0 forms a new, more powerful approach in the assessment of and treatment planning for individuals presenting with hypernasality.
References


DANKWOORD

Tijdens het schrijven van mijn doctoraat ging ik op zoek naar een definitie van het begrip resonantie. Ik vond het volgende: “Resonantie ontstaat wanneer een trillend object zijn trillingen overbrengt op een ander object en dit ander object in het ritme van de oorspronkelijke trillingen gaat meetrillen. Hierbij kan de trilling van dit andere object veel sterker zijn dan men op grond van de aanstoting kan verwachten.” Bij het lezen van deze definitie besefte ik dat resonantie niet alleen centraal staat in mijn doctoraat, maar dat het ook vaak optrad tijdens de weg die ik ondertussen heb afgelegd. Er zijn namelijk heel wat mensen die hun enthousiasme, interesse, aanmoedigingen en steun overgebracht hebben op mij en er zo voor gezorgd hebben dat ik op het juiste moment een impuls kreeg om mijn beweging te starten, verder te zetten of te versterken. Ik zou hen hier dan ook graag voor willen bedanken.

Eerst en vooral wil ik mijn promotor, prof. dr. Kristiane Van Lierde, bedanken. Kristiane, u gaf mij de eerste impuls tijdens mijn doctoraatstraject door in mij te blijven geloven en mij als assistent een plaats te geven in uw team. Een welgemeende dank u wel voor uw begeleiding, enthousiasme, en aanmoediging, om mij af en toe dat duwtje te geven dat mij dwong om buiten mijn comfortzone te treden, om uw kennis over resonantiestoornissen en tal van andere (logopedische) topics met mij te delen en mijn vaardigheden als wetenschapper, clinicus en lesgever te helpen ontwikkelen.

Dit werk was ook niet tot stand kunnen komen zonder de significante impulsen van mijn copromotor, prof. dr. Floris Wuyts. Floris, u liet mij de soms duistere wereld van de statistiek ontdekken. Bedankt voor uw kritische ingesteldheid en creatieve impulsen. En hoewel de naam van de index uiteindelijk ‘NSI 2.0’ is geworden, weten alleen u en ik de hoeveelste versie van de NSI het werkelijk is.


Een extra dank u wel voor prof. dr. Paul Corthals, voor uw geduldige hulp bij de akoestische analyses en het uitpluizen van de VLHR. En voor prof. dr. Marc De Bodt en dr. Youri Maryn voor de intensieve beoordelingssessies. Bedankt ook, alle leden van het schisisteam, om mij als collega op te nemen in jullie team en jullie ervaring met mij te delen.
Dankwoord

Een bijzonder woord van dank richt ik graag aan prof. dr. Hubert Vermeersch. U liet mij een nieuwe wereld ontdekken door mij de kans te geven mee te gaan naar Oeganda en deel uit te maken van het VLIR-UOS project. Tijdens de missies, die mij niet alleen professioneel, maar ook persoonlijk bewogen hebben, mocht ik vele warme en inspirerende personen ontmoeten. Bedankt hiervoor.

Zonder de input van alle kinderen, hun ouders en de vele controlepersonen die deelnamen aan mijn onderzoek zou ik al snel stilgevallen zijn. Bedankt voor jullie enthousiasme, extra bezoekjes aan het UZ en hartelijke ontvangst. Ook bedankt aan de studenten die ik mocht begeleiden tijdens hun masterproef en die zo voor extra impulsen zorgden.


Ook alle collega’s van het UZ zorgden voor de nodige ontspanning tijdens de middagpauzes, al dan niet in de zon, of de vele treinritten. Na zo’n adempauze was er weer energie om mijn beweging verder te zetten. Dankjewel.

Lieve Sofie, hoe zou ik deze soms helse periode zonder jou overleefd hebben? Sinds dag één brainstormen we samen over de obstakels die we tegenkomen, vaak met een impuls vanuit een andere richting als resultaat of een nieuw inzicht gewoon door het even te kunnen vertellen. Als de stoom uit mijn oren kwam, maar ik weigerde om iets te zeggen, bracht jij raad en rust. En als dat niet voldoende was, was er nog altijd tequila… Je eindeloze hulpvaardigheid, je bezorgdheid en je warme persoonlijkheid maken je tot iemand die ondertussen meer geworden is dan een collega of bureaumaatje. Bedankt!

Daarnaast heb ik het geluk om enkele trouwe supporters te hebben die al bijna mijn hele leven op dezelfde frequentie zitten als ik. Bedankt Julie, Evelyn, Dora, Caroline, Wim, Tineke, Eline, Elisabeth, Glenn, Elke en Matthias voor jullie luisterend oor, oprechte interesse, gezellige etentjes, zotte feestjes en markante avonden. We bouwen allemaal aan onze eigen weg, maar het is fijn om te weten dat ik altijd terecht kan bij mensen die mij zo goed kennen.

De warme omgeving waarin ik mocht opgroeien en waarin ik mezelf mag zijn, heb ik te danken aan een geweldige familie en schoonfamilie. De manier waarop iedereen meegeleefd heeft tijdens deze 5 jaar heeft mij enorm veel steun gegeven. Astrid, Elien en Bart de gezellig drukke etentjes op vrijdagavond zijn nog steeds iets waar ik elke week weer naar uitkijk. Oma, de rots van de familie en zo ongelofelijk trots op al haar kleinkinderen.
**Dankwoord**

Mireille en Martin, bedankt voor de wekelijkse etentjes, de schouderklopjes en zorgzame aandacht.

Mama en papa, jullie zijn de trotse bron van mijn beweging. Jullie hebben mij steeds de vrijheid gegeven om mijn eigen keuzes te maken, af en toe bijgestaan door wat goede raad, maar altijd met jullie volledige steun. Bedankt voor al jullie hulp, liefde en warme gezelligheid. Bedankt om er mee voor te zorgen dat ik vandaag sta waar ik nu sta.

Lieve Fréderic, mijn laatste woorden heb ik voor jou bewaard. Ik ben je dan ook een hele lieve, warme en dikke dankjewel verschuldigd. Je bent die ene persoon die mij uitdaagt om mijn grenzen te verleggen, die elke dag klaar staat voor mij, zelfs als het mij even te veel wordt, die onvoorwaardelijk mijn soms ingewikkeldte handleiding blijft uitpluizen en die mij soms beter kent dan ik mezelf ken. Je haarfijne analyses bieden mij vaak een perspectief waaruit ik opnieuw kan groeien. Bedankt om mij zo ongelooflijk graag te zien.