

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/105347>

Please be advised that this information was generated on 2019-04-21 and may be subject to change.

Use of a Perceptual Evaluation Instrument to Assess the Effects of Infant Orthopedics on the Speech of Toddlers With Cleft Lip and Palate

EMMY M. KONST, PH.D.
TONI RIETVELD, PH.D.
HERMAN F.M. PETERS, PH.D.
HANNY WEERSINK-BRAKS, M.A.

Objective: To investigate the effects of infant orthopedics (IO) administered in the first year of life on the speech characteristics of 2.5-year-old children with complete unilateral cleft lip and palate (UCLP) using a perceptual evaluation instrument with equal-appearing interval (EAI) scales.

Design: In a prospective randomized clinical trial (Dutchcleft), two groups of children with complete UCLP were followed longitudinally. One group received IO based on a modified Zurich approach (IO group), and the other group did not (non-IO group). The appliance was used until soft palate closure at age 12 months. Hard palate closure is delayed until 9 years of age.

Participants: Three groups of 2.5-year-old toddlers participated in this investigation: 10 IO, 10 non-IO, and 8 noncleft controls matched for age and socioeconomic status.

Method: Five trained listeners assessed the children's speech in a blinded perceptual rating procedure. They judged 13 specific speech characteristics and indicated their total impression of speech on EAI scales.

Results: The reliability and consistency of 11 of the rating scales was good. The intelligibility rating scale was the single speech characteristic that distinguished the IO group from the non-IO group; the IO group was judged to be superior. The cleft groups differed from the noncleft group on 9 of the 11 scales.

Conclusions: Evaluation of speech by means of the present newly developed perceptual rating instrument showed that the IO group obtained significantly higher ratings for intelligibility than the non-IO group. The groups did not differ regarding any of the other speech aspects.

KEY WORDS: *cleft lip and palate, infant orthopedics, perceptual rating scale, randomized prospective clinical trial*

In many cleft palate centers, especially in Europe, babies with cleft lip and palate are treated with infant orthopedics (IO). There is considerable controversy regarding this treatment. Proponents claim that IO provides correct alignment of the alveolar segments and narrowing of the cleft (McNeil, 1956), facilitation of surgical closure of the cleft and thus improved aesthetic outcome (Gnoinski, 1990), facilitation of feeding (Oliver, 1973; Lubit, 1976), parental support (Lubit, 1976; Huddart, 1990), and improved speech (Stuffins, 1981;

Hotz et al., 1986; Gnoinski, 1990). Opponents state that this treatment is expensive, inhibits maxillary growth (Pruzansky, 1964), and increases the incidence of dental caries (Bokhout et al., 1996). These reports are largely anecdotal and based on clinical observations, except for the study by Bokhout et al. (1996). Therefore, these studies do not provide convincing, scientifically documented arguments regarding the value of IO. To investigate the effects of treatment with IO based on a modified Zurich approach, a three-center, randomized, prospective clinical trial (Dutchcleft) was started in 1993 in the Netherlands.

Earlier reports on the results of the Dutchcleft trial suggest that IO administered in the first year of life is beneficial to the speech and language development in the short term as well as the long term. At 12 months of age and with the plate still being *in situ*, children treated with IO showed enhanced production of alveolar contoids in babbling (Konst et al., 1999). Their phonological development from age 2 to 3 years was closer to normal when compared with the non-IO group (Konst

Dr. Konst is a Speech-Language Pathologist in the Department of Orthodontics and Oral Biology, and Dr. Peters is Associate Professor of Speech Pathology and Dr. Weersink-Braks is a Speech-Language Pathologist formerly affiliated with the Department of Voice and Speech Pathology, University Medical Centre St. Radboud, University of Nijmegen, the Netherlands. Dr. Rietveld is Associate Professor of Phonetics, Department of Language and Speech, University of Nijmegen, the Netherlands.

Submitted February 2002; Accepted January 2003.

Address correspondence to: Dr. Emmy M. Konst, Department of Orthodontics and Oral Biology, University of Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, the Netherlands. E-mail e.konst@dent.umcn.nl.

et al., 2003b). At age 2.5 and 3 years, children in the IO group produced more complex sentences with a larger mean length of utterance (MLU) (Konst et al., 2003a). Furthermore, at the age of 2.5 years, the IO group received higher ratings for speech intelligibility (Konst et al., 2000). It is plausible that the beneficial effect of IO on speech is also reflected in other, more specific speech characteristics such as palatalization or lateralization. The present article describes a perceptual evaluation of the speech of 2.5-year-old children carried out to evaluate effects of IO on speech characteristics.

Perceptual evaluation via rating scales is a frequently applied and potentially useful technique to evaluate speech characteristics. Several procedures for assessing speech via rating scales exist. In their review, Kreiman et al. (1993) pointed out that equal-appearing interval (EAI) scales, visual analog scales, direct magnitude estimation (DME), and paired comparison tasks have all been employed as perceptual evaluation procedures to evaluate speech or voice quality. Of these methods, EAI scaling (usually with seven-point scales) has been applied most frequently. Although many studies have been published using perceptual evaluation procedures to assess cleft palate speech, it is unclear which of the many rating scales, procedures, and statistics are best suited to measure characteristics of cleft palate speech. Particularly for this patient group, for which development of normal speech is one of the most important goals, the need for an adequate and reliable perceptual evaluation instrument is urgent. Initiatives have already been undertaken to standardize the parameters of speech evaluation in the clinical setting (Sell et al., 1994; Wyatt et al., 1996; Hirschberg and Van Demark, 1997; Shaw et al., 2000; Sell et al., 2001). A framework for perceptual evaluation of cleft palate speech in a multilingual perspective was provided by the Eurocleft Speech Group (1993). However, there is still considerable controversy about which parameters of speech should be evaluated and how they should be measured.

In this article, a specific set of perceptual rating scales was developed to assess the impact of IO on the speech characteristics of 2.5-year-old toddlers with cleft palate. In addition, the rating scales were used to compare the speech of these children with a noncleft control group. Only those rating scales that were rated reliably and consistently were used to evaluate the speech.

METHOD

Study Design and Treatment Protocol

The study was designed as a three-center, prospective, two-arm, randomized, controlled clinical trial. The infants who were included had complete unilateral cleft lip and palate (UCLP) and were born at full term. Their parents were all Caucasian, and Dutch was their native language. Patient exclusion criteria included the presence of other congenital malformations (except for syndactyly) or soft tissue bands. Parents of eligible infants were verbally informed about the trial. The

formal informed consent contained comparable written information and was signed by the parents after they had agreed to allow their children to participate. All infants entered the trial within 2 weeks after birth and were randomly assigned to one of two groups: the IO group (which received IO in the first year of life) or the non-IO group (which did not receive such treatment). All other interventions were the same in both groups.

IO treatment thus started within 2 weeks after birth. The technique used was described earlier (Prahl et al., 2001). The appliance, made of soft and hard acrylic, was worn 24 hours a day until soft palate closure. The mean duration of IO was 50 weeks. At 18 weeks, the lip was closed surgically according to the Millard technique. The palate was closed in two stages, with soft palate closure (modified Von Langenbeck procedure) at 12 months of age. Hard palate closure is delayed until approximately 9 years of age. To standardize the treatment, all specialists participated in consensus meetings.

Participants

A total of 54 babies (41 boys, 13 girls) was included in the trial, 27 in the IO group and 27 in the non-IO group. In total, the intake of this patient sample covered a period of more than 3 years. A detailed description of the sample characteristics was given by Prahl et al. (2001). The data in this article were derived from three groups of 2.5-year-old toddlers: a group of 10 children with UCLP treated with IO (eight boys, two girls; age range 2 years, 5 months, 24 days to 2 years, 6 months, 9 days), a group of 10 children with UCLP treated without IO (nine boys, one girl; age range 2 years, 5 months, 15 days to 2 years, 6 months, 8 days), and a control group of eight noncleft peers (two boys, six girls; age range 2 years, 6 months, 8 days to 2 years, 7 months, 5 days). The noncleft controls were matched for age and socioeconomic status. The sample assessed in this article is smaller than the total number of children participating in the full trial because of the lengthy recruitment period. At the time of the present evaluation, only 10 children with UCLP of this age were available in each group.

Data Collection

A sample of spontaneous speech was recorded from all the participants in the child's home environment by two investigators. The speech was recorded using high-quality audio equipment (TCD-D7 DAT Walkman with a Sennheiser MD421U-4 dynamic microphone; Sony, Tokyo, Japan) while the child was engaged in a semistructured play scenario with one of the researchers. A fixed set of toys likely to elicit words that are most commonly heard in the active vocabulary of the normally developing 2.5-year-old child was used (Schlichting et al., 1995).

Rating Scales

The rating scales employed were selected from the literature to constitute a comprehensive perceptual instrument for the evaluation of the results of IO treatment on the speech characteristics of children with UCLP. As a starting point, the phonetic abnormalities associated with cleft palate that are mentioned in the literature were taken. Among the abnormalities described in the literature are weakened high-pressure consonants, audible nasal emission, palatalization of consonants, backing of consonants, frequent use of glottal stops (Stengelhofen, 1989), misarticulation of /s/, nasopharyngeal snort (Morley, 1970), aspirate phonation, and glottal stop and pharyngeal fricative compensatory articulations (Bzoch, 1979). Additionally, hypernasality and dysphonia are also mentioned (McWilliams et al., 1984). Furthermore, speech parameters proposed by others were also taken into account when selecting the rating scales to be used in this evaluation (Dalston et al., 1988; Eurocleft speech group, 1993; Sell et al., 1994, 2001; Harding and Grunwell, 1996; Hirschberg and Van Demark, 1997). The perceptual evaluation instrument comprised the following rating scales.

Place of Articulation

1. Palatalization. On this scale, a speech characteristic is rated that may occur in alveolar and velar consonants. When palatalization occurs, the target consonant is produced with palatal articulation (i.e., the blade of the tongue makes contact with the hard palate).
2. Lateralization. Scores on this scale indicate whether the voiceless fricative /s/ is realized with lateral airflow rather than with central airflow.
3. Fronting. Ratings on this scale specify whether an alveolar consonant is produced with a more fronted place of articulation (i.e., dental or interdental).
4. Backing. This rating scale applies to the articulation of both plosive and fricative consonants. If the target consonant is produced with a more posterior place of articulation resulting in a phonetic change in the sound, the item should be awarded a low score. If the backing results in a phonemic change, a high scale value should be given.
5. Glottal Articulation. This aspect refers to both the occurrence of a glottal stop simultaneously with the articulation of a plosive or fricative and a substitution of these consonants by a glottal stop. Glottal stops are used to compensate for the fact that the air pressure built up in the oral cavity during the production of pressure consonants is insufficient.

Voice Characteristics

1. Hyperkinetic Voice. This rating scale was included because a hyperkinetic voice may also relate to difficulties in oral pressure build-up. It is characterized by strong contraction of the muscles participating in phonation and results in a harsh and strained voice.

Nasalization

1. Hypernasality. Hypernasality in this experiment was defined as excessive nasal resonance accompanying normally nonnasalized target vowels. This speech aspect is associated with velopharyngeal insufficiency.
2. Nasal Emission. On this scale, the incidence of audible escape of nasal airflow during the production of a consonant is rated. This speech aspect is associated with velopharyngeal insufficiency.
3. Nasal Fricative. This characteristic receives a high score if a turbulent airflow in the nasal cavity occurs simultaneously with a fricative or as a substitution for a fricative.
4. Nasal Snort. Nasal snort is defined as velopharyngeal friction during the production of a consonant. This characteristic applies to both plosives and fricatives.
5. Nasal Realization. If an oral consonant is substituted by a nasal consonant, this is scored as nasal realization.

General Evaluative Characteristics

1. Correctness of Articulation. This rating scale refers to the extent to which the articulation of an utterance is correct.
2. Intelligibility. On this scale, the listeners rated how much of the utterance they were able to understand.
3. Total Impression. The listeners were asked to consider all the above aspects of speech quality to determine their overall impression of the child's speech. This scale was included to obtain a general, summarizing effect measure to be used in the cost-effectiveness analysis that was part of the clinical trial (Severens et al., 1998).

All EAI scales, except for the scale "total impression," were seven-point rating scales. Seven-point scales are most frequently used in the literature. In practice, most listeners do not use the whole range because they avoid scale extremes (Kreiman et al., 1993). A seven-point scale offers the listener enough detail and discriminative power to distinguish various levels of severity. A scale with 10 points was used for the listeners' total impression of the speech. This score range, which corresponds with the grades given in Dutch schools, was considered appropriate for the scale "overall impression" because we explicitly asked the listeners to give the child a report mark for his/her speech.

The scales were constructed in such a way that a high scale point indicates that a speech characteristic is strongly present. On the scales "intelligibility," "correctness of articulation," and "overall impression," a high scale value corresponds with better speech. Because all the other scales represent speech errors, a high scale value on these scales corresponds with more disordered speech.

Procedure

From the recorded conversations with the children, a speech sample of 15 utterances per child was selected for use in the

blinded perceptual evaluation. All these utterances were spontaneous with a clarified exact meaning for all words. The ratings were carried out by five trained female listeners. All were graduated speech therapists with experience in assessing cleft palate speech, but they were not involved in the children's treatment.

Before the actual experiment took place, the raters attended a specially designed training session allowing them to familiarize themselves with the scales and speech samples. First, the listeners were asked to absorb the definition of a particular scale after which they listened to two typical examples of the speech characteristic involved (anchor stimuli). In one example the speech characteristic was strongly present, in the other example less so. Second, the listeners were asked to rate two samples of cleft palate speech (not included in the experiment) to practice the specific demands of the rating task.

The actual rating procedure took approximately 2 hours and comprised a total of 34 listening blocks (i.e., 20 listening blocks of speech samples of the children with UCLP [10 IO and 10 non-IO], eight samples from the typically developing children, and six listening blocks that were presented twice [not consecutively] for calculation of intrarater reliability). The order in which the listening blocks were presented to the listeners was randomized. Each listening block included four repetitions of the same speech sample. The MLU of the speech sample equaled the MLU of the entire recorded conversation. The orthographic transcription of each speech sample was printed on the scoring sheet. The listeners were asked to read the transcription after which they were presented with three repetitions of the speech sample allowing them to make all necessary ratings. The time between each of the three presentations of the speech sample was 5 seconds. After the third presentation, a pause of 10 seconds was given to allow the listeners to finish their specific judgments and turn the page. The fourth presentation of the speech sample afforded them the opportunity to give their total impression of the speech. Pauses of 20 seconds were included between the listening blocks. An example of the scoring sheet is given in the Appendix.

Statistical Analysis

The first step in the analysis procedure was to determine the reliability of the measuring instrument. The concept of interrater reliability implies that the listeners rate the objects in a parallel fashion, although the absolute rating may vary from listener to listener. The interrater reliability of the ratings on the interval scales was calculated by means of Cronbach's α , the appropriate statistic when raters are considered as a fixed factor, and the objects as a random factor (Rietveld and Van Hout, 1993). A scale's reliability was considered sufficient if Cronbach's α exceeded 0.75 (Van Erp, 1991). Intrarater reliability was assessed by means of Pearson's correlation coefficient between ratings and reratings. The consistency of the ratings was evaluated by *t* tests, which compared the absolute values of the ratings and reratings.

TABLE 1 Interrater Reliability (Cronbach's α) and Intrarater Reliability (Pearson's *r*) of the Rating Scales

Rating Scale	Cronbach's α	Pearson's <i>r</i>
Palatalization	0.81	0.85*
Lateralization of /s/	0.84	0.81*
Fronting	0.23	
Backing	0.95	0.98**
Glottal articulation	0.90	0.96**
Hyperkinetic voice	0.78	0.89*
Hypernasality	0.90	0.94**
Nasal emission	0.90	0.98**
Nasal fricative	0.90	0.95**
Nasal snort	0.87	0.63 ns†
Nasal realization	0.89	0.98**
Correctness of articulation	0.93	0.99**
Intelligibility	0.96	0.99**
Total impression	0.98	0.98**

† ns = not significant.

* Pearson correlation significant at $p < .05$.

** Pearson correlation significant at $p < .01$.

Second, the ratings on the scales that were rated reliably and consistently were used to identify differences in speech characteristics among the IO, non-IO, and noncleft control groups. To this end, series of univariate analyses of variance (ANOVA) were carried out followed by *post hoc* tests (Tukey's honest significant difference [HSD] test) when the omnibus *F* ratio was statistically significant.

For the scales that showed a statistically significant difference between the IO and non-IO group, the effect size was calculated. Effect sizes may be used to assess the magnitude and meaning (clinical importance) of a treatment effect (Kazis et al., 1989). The effect size in this study was expressed as $ES = (m_{IO} - m_{non-IO}) / \text{overall SD}$, where m_{IO} is the mean rating in the treated group (IO), m_{non-IO} is the mean rating in the non-IO group, and overall SD is the SD pooled for both groups (Cohen, 1977). Cohen defined an effect size of 0.20 as small, one of 0.50 as moderate, and one of 0.80 or greater as large.

RESULTS

Interrater Reliability

The reliability of the ratings was assessed by computing Cronbach's α for each rating scale. The scales and corresponding reliability coefficients are listed in Table 1. When Cronbach's α exceeded 0.75, scales were considered to be scored reliably, which was the case for all scales except for the scale "fronting." The results on this scale were therefore excluded from further analysis.

Intrarater Reliability

Intrarater reliability was calculated by means of Pearson's correlation coefficient. Because a high correlation between a test and retest does not guarantee that the objects are given the same ratings in both tests, this analysis was followed by *t* tests to determine whether the values of the test and retest ratings differed significantly in magnitude. Table 1 shows the corre-

TABLE 2 Results of the ANOVA Analyses: Mean Scores and Standard Deviations (in parentheses) and Significance of Group Differences on 11 Rating Scales for 10 Children With Unilateral Cleft Lip and Palate (UCLP) With Infant Orthopedics (IO), 10 Children With UCLP Without IO (nIO), and Eight Children Without Cleft (nclp)

Rating Scale	IO (n = 10)	nIO (n = 10)	nclp (n = 8)	F Ratio	Contrast Between***
Palatalization	2.28 (1.26)	2.28 (1.20)	1.03 (1.33)	$F_{(2,25)} = 2.84$	
Lateralization of /s/	1.98 (1.69)	1.60 (1.23)	1.30 (1.16)	$F_{(2,25)} = 0.54$	
Backing	4.08 (1.79)	4.08 (1.79)	1.18 (1.22)	$F_{(2,25)} = 13.55^{**}$	IO-nclp nIO-nclp
Glottal articulation	1.94 (1.35)	2.12 (1.35)	0.23 (0.33)	$F_{(2,25)} = 6.99^{**}$	IO-nclp nIO-nclp
Hyperkinetic voice	1.54 (1.02)	1.42 (0.82)	0.60 (0.24)	$F_{(2,25)} = 3.56^*$	IO-nclp
Hypernasality	3.28 (1.08)	3.66 (0.98)	0.58 (0.39)	$F_{(2,25)} = 30.30^{**}$	IO-nclp nIO-nclp
Nasal emission	2.76 (1.17)	3.76 (1.15)	0.38 (0.42)	$F_{(2,25)} = 25.74^{**}$	IO-nclp nIO-nclp
Nasal fricative	3.06 (1.47)	3.92 (1.59)	0.45 (0.50)	$F_{(2,25)} = 16.11^{**}$	IO-nclp nIO-nclp
Correctness of articulation	1.88 (1.01)	1.24 (0.48)	4.15 (0.95)	$F_{(2,25)} = 7.35^{**}$	IO-nclp nIO-nclp
Intelligibility	2.62 (1.32)	1.44 (0.65)	4.78 (0.78)	$F_{(2,25)} = 28.71^{**}$	IO-nIO IO-nclp nIO-nclp
Total impression of speech quality	3.52 (1.75)	2.18 (0.62)	7.13 (1.49)	$F_{(2,25)} = 26.43^{**}$	IO-nclp nIO-nclp

* Significant at $p < .05$.** Significant at $p < .01$.*** Contrast between groups significant at $p < .05$ in Tukey's HSD test.

lations between the test and retest samples for each speech scale. Intrarater reliability was high for all scales with the exception of the characteristic "nasal snort," signifying that, apart from this aspect, the listeners were consistent in their ratings. The characteristic "nasal snort" was subsequently excluded from further analysis.

All t tests, apart from the one for "nasal realization," were not statistically significant at the 5% level. The t test for the ratings on "nasal realization" was significant at $p = .04$ ($t_5 = 2.7$); the ratings in the first judgment were on average 0.3 scale points higher than those in the retest. Because ratings on this scale had not been consistent, they were excluded from further analysis.

Use of the Rating Instrument in Evaluating the Effects of IO

As described in the previous section, the three aspects that were not rated reliably or consistently (i.e., "fronting," "nasal snort," and "nasal realization") were excluded from further analyses. The remaining 11 reliably scored rating scales were used to evaluate the effects of treatment with IO and compare the children with cleft lip and palate to the control group of noncleft peers. To this end, univariate ANOVAs followed by *post hoc* comparisons (Tukey's HSD test) were carried out for each speech variable. Prior to this analysis, the homogeneity of the variances was tested by means of Levene's test of homogeneity. None of the scales yielded significant data, indicating that the homoscedasticity assumption was not violated

and therefore univariate ANOVA was appropriate. The results of these tests are listed in Table 2. The two cleft groups differed from the noncleft group on most rating scales, with the characteristics palatalization and lateralization being the two exceptions. On these scales, the three groups obtained similar ratings. As far as differences between the two cleft groups are concerned, the scale "intelligibility" was the only rating scale that could discriminate between the IO and non-IO groups. The IO group obtained significantly higher intelligibility ratings than the non-IO group, indicating that the intelligibility of the IO group's speech was superior. Higher ratings on the other scales implied that these speech characteristics were judged to be more distinctly present in the speech sample. As expected, the cleft-related speech aspects were judged to be least present in the noncleft control group. Also, the intelligibility and correctness of articulation ratings were highest for this group.

Effect Size

A statistically significant effect of treatment with IO was found on the intelligibility scale. To evaluate the meaning of this statistically significant difference, the effect size $ES = (m_{IO} - m_{non-IO})/overall\ SD$ was calculated. ES yielded 1.0 $((2.62 - 1.44)/1.17)$, which is large.

Correlation Among the Ratings

The rating scale "total impression of speech quality" was included to obtain a general, evaluative measure that could be

TABLE 3 Results of the Multiple Regression Analysis With Intelligibility as Dependent Variable

Predictor Variable	Beta	Significance	Cumulative R ²
Correctness of articulation	0.75	$p < .001$	0.89
Lateralization	0.21	$p < .001$	0.91
Backing	-0.28	$p < .01$	0.93

used in the cost-effectiveness analysis that was part of the clinical trial (Severens et al., 1998). Pearson correlations were calculated to determine the relationship of this general variable with each of the other, more specific variables. As was to be expected, the variables intelligibility ($r = 0.97$) and correctness of articulation ($r = 0.96$) both correlated highly with total impression of speech quality. There was also a negative high correlation with backing ($r = -0.85$) and hypernasality ($r = -0.82$), indicating that the presence of the characteristics backing and hypernasality correlated with lower ratings for total impression. All correlations were significant at $p < .001$.

In this study, intelligibility was considered a key outcome variable, not only because this was the single speech aspect that distinguished the IO group from the non-IO group but also because this speech characteristic was regarded as a crucial facet of communication. To gain more insight into the speech aspects that had contributed to the intelligibility ratings, a multiple regression analysis (stepwise procedure) was performed with intelligibility as the dependent variable (Table 3). The number of predictor variables in the regression analysis was nine. Three of the predictor variables were included in the regression equation. These variables, correctness of articulation, lateralization, and backing, explained 93% of the variance of the intelligibility rating. There was a positive correlation between intelligibility and correctness of articulation and between intelligibility and lateralization and a negative correlation between intelligibility and backing.

DISCUSSION

The effects of treatment with IO in the first year of life on later speech of toddlers with UCLP were evaluated by means of perceptual ratings on a set of EAI scales. EAI scaling is often used in circumstances in which no objective measuring instrument is available, but it has some limitations. The validity of a measurement procedure rests on the psychophysical nature of the dimension that is studied. Stevens (1975) indicated that dimensions are either prothetic or metathetic in nature. According to Stevens, a prothetic continuum is additive, whereas a metathetic continuum is substitutive. An example of a prothetic continuum is loudness. This aspect of speech varies in magnitude or quantity. A prothetic continuum such as loudness cannot be partitioned into equal intervals by listeners. If raters attempt to partition a prothetic dimension into equal intervals, they typically demonstrate a systematic bias toward subdividing the lower end of the continuum into smaller intervals than the upper portion of the continuum. Interval scal-

ing, therefore is said to represent measurement of a prothetic continuum poorly. DME scaling does not assume a linear partitioning of the continuum and is a more appropriate procedure for prothetic dimensions.

An example of a metathetic continuum is pitch. With a metathetic dimension, listeners are able to divide the continuum into equal intervals because the listeners' naturally occurring perceptual intervals are equal. Therefore, EAI scales are suitable for quantification of a metathetic dimension.

Many of the perceptual dimensions commonly scaled in speech pathology, such as nasal voice quality or intelligibility, are prothetic in nature (Schiavetti, 1992; Zraick and Liss, 2000) and may therefore be poorly represented by interval scaling. Nevertheless, many investigators have used EAI scaling to measure both metathetic and prothetic dimensions because DME may not be the most practical method for clinical measurement of speech. For example, in measuring intelligibility, the lack of an easily interpreted unit of measurement reduces the clinical utility of DME for communication of intelligibility data to other professionals or laypersons. Additionally, the DME procedure can be somewhat cumbersome to use because it requires the use of either a standard speech sample assigned to a modulus value or the difficult modulus equalization technique to remove interlistener variance in selection of a modulus value for the free-modulus procedure (Schiavetti, 1992). Folkins and Moon (1990) furthermore, indicate that the differences between the EAI and DME scaling may not always justify the extra work required to perform DME procedures. In the present study, EAI scaling was preferred over DME for practical reasons. First of all, the DME approach requires a reference stimulus for each speech dimension that is measured. In a multidimensional approach as in the present experiment, listeners rate all dimensions one after another per listening block. This procedure implies that if a reference stimulus is used, this stimulus should correspond to the same modulus value on the scale for all dimensions. Obviously, this is impracticable. Another reason for using EAI scaling in this study was that our expert raters were familiar with this type of scaling. The use of a new method would probably have introduced additional interrater variability, thereby reducing the reliability of the data. It should be noted, however, that in perceptual rating experiments evaluating one single dimension, the use of a DME approach may prevail over EAI scaling because the power of the scale may be enhanced in a DME approach.

The actual rating experiment in our study was preceded by a training in which so-called anchor stimuli were used. For each rating scale, two anchor stimuli with the corresponding scale value were displayed to the assessors. The scale values of the anchors had been determined beforehand through consensus listening performed by another panel of expert listeners. The training was included because it has been reported that even highly experienced listeners frequently completely disagree on what they hear (Kreiman et al., 1993). Training with reference samples, as was done in our study, may promote

better interrater agreement than simply selecting experienced raters (Kent, 1996).

Before any ratings could be used in the evaluation of IO treatment, we needed to establish whether our human measuring instrument was indeed reliable and consistent. If, like in many clinical research settings, we had employed a single rater or consensus rating by more than one rater, the reliability of the ratings could not have been determined. We therefore opted for a panel of listeners. The interrater reliability (Cronbach's α) and intrarater reliability (Pearson's r) of the ratings were calculated for all interval rating scales. There was one scale on which the judgments were not reliable: the scale "fronting." The difficulty that raters experienced in judging "fronting of speech" may be explained by the fact that this speech aspect is difficult to perceive from auditory stimuli only. Typically, visual information is used as well in rating this feature.

There were two other scales, viz. "nasal snort" and "nasal realization," that were not scored consistently. The Pearson correlation between the test-retest samples on the scale "nasal snort" was not statistically significant ($r = 0.63$; $p > .05$); therefore, it was decided to exclude the ratings on this scale. On the scale "nasal realization," the magnitude of the ratings in the retest was 0.3 scale point lower than in the first rating ($t_5 = 2.7$; $p < .05$). The ratings on this scale were therefore considered inconsistently scored, and this speech characteristic was subsequently excluded from the analysis.

The scales that were scored reliably in this experiment proved an adequate instrument by means of which the cleft groups could be distinguished from the noncleft control group, with the exception of the speech aspects lateralization and palatalization. On the latter scales, all three groups obtained similar scores. This can be explained by the fact that these errors in the placement of articulation are not uniquely confined to the cleft palate population. Indeed, these errors were present in one of the noncleft controls who received high ratings for palatalization and lateralization. All other reliably rated speech characteristics in this experiment distinguished the cleft groups from the control group. All speech errors were scored as being most distinct in the two toddler groups with cleft; the intelligibility and correctness of articulation was highest in the children without cleft. When comparing the ratings for the cleft groups with the noncleft control group, it should be noted that the noncleft control group was matched for age and socioeconomic status but not for MLU. The expressive skills of the noncleft control group, therefore, may be expected to be more mature than those of the cleft groups.

The present rating experiment was carried out to investigate the effects of the use of IO in the first year of life on the speech proficiency of children with UCLP at age 2.5 years. We found a positive effect of IO on intelligibility only. The difference between the IO and non-IO group (1.18 scale point) was statistically significant. The effect size showed that this effect of IO treatment was large, indicating that it may be considered as clinically relevant. A beneficial effect of IO on the judged speech intelligibility was also demonstrated in another rating experiment with lay listeners (Konst et al., 2000)

who assessed the intelligibility of speech samples of the same group of toddlers. Konst et al. (2000) also described an intelligibility assessment in which listeners transcribed the utterances in a write-down procedure. This write-down measurement did not present statistically significant effects of IO treatment on intelligibility.

The explanation for enhanced speech intelligibility after IO treatment is not by definition transparent. IO is used only in the first 12 months of life, but it appears to have longer-term effects. This long-term effect of IO can be explained by its impact on the oral motor patterns that are established in the first year. It is known that a cleft palate already influences the development of the movements of the tongue tip in the very early months of life (Morley, 1970). Many children with alveolar clefts may develop greater activity of the mid and posterior parts of the tongue blade in sucking because the tongue is unable to compress the nipple against the alveolus. The increased activity of the mid and posterior parts of the tongue is also seen in the development of speech; because there is no support for the tongue, alveolar articulations will be produced at a more posterior place of articulation. Treatment with IO provides the infant with an artificial alveolar ridge and may thus aid the development of more normal sensorimotor patterns and minimize compensatory behavior in feeding and babbling. Results on the short-term effects of this longitudinal clinical trial have shown that children with IO produce more alveolar articulations than non-IO children in the babbling period, when the plate is still *in situ* (Konst et al., 1999). The use of alveolar sounds is predominant in the babbling of infants without cleft (Smith and Oller, 1981; Stoel-Gammon, 1985; Smith, 1988) and less present in the babbling of children with cleft (Grunwell and Russell, 1987; O'Gara and Logemann, 1988; Estrem and Broen, 1989). The use of alveolar consonant sounds in the prelexical period may reflect the establishment of more normal sensorimotor speech patterns in children treated with IO. It is likely that these children develop less deviant articulatory movements, which may result in more intelligible speech.

Multiple regression analysis demonstrated that the judged intelligibility correlated positively with correctness of articulation and lateralization and negatively with backing. The positive correlation with correctness of articulation is obvious, as is the fact that less retracted speech is judged to be more intelligible. The positive correlation between lateralization and intelligibility, indicating that lateralization of speech was associated with better intelligibility, however, is less clear. It may relate to two phenomena in the data. First of all, the speech of the children with cleft palate who obtained high scores for intelligibility was almost free of speech errors except for the feature lateralization. Second, in the noncleft control group, which received the highest intelligibility ratings, lateralization was almost as strongly present as in the two cleft palate groups. Therefore, although the connection is not causal, in this experiment lateralization positively correlates with intelligibility. The judged intelligibility also correlated highly with the ratings on the total impression of speech quality. The same

correlation was found by Preminger and Van Tasell (1995, p. 722) who concluded that “overall impression of speech quality can be predicted with some confidence on the basis of judged intelligibility when the conditions are such that intelligibility varies widely.”

CONCLUSIONS

Evaluation of speech of 2.5-year-old children with complete UCLP by means of a newly developed perceptual rating instrument showed that IO treatment during the first 12 months of life affected the judged intelligibility. Children who were treated with IO in the first year of life obtained higher ratings for intelligibility than the non-IO children. The perceptual evaluation instrument comprised 14 EAI rating scales of which 11 were scored reliably and consistently. These 11 scales proved an adequate instrument for assessing speech in children with cleft lip and palate.

Acknowledgments. This study was supported by a grant from the National Health Insurance Board of the Netherlands and is part of the Dutch three-center research project (called Dutchcleft) into the effects of IO treatment in complete UCLP. The project is carried out at the Cleft Palate Centre of the University of Nijmegen in close collaboration with the Cleft Palate Centers of the Free University of Amsterdam and the Erasmus University of Rotterdam (coordinating orthodontists A. M. Kuijpers-Jagtman and B. Prah-Andersen). The authors express their gratitude to the Cleft Palate Centers of Nijmegen, Amsterdam, and Rotterdam with special thanks to the speech pathologists and the ear-nose-throat specialists. We are also indebted to K. Jansonius-Schultheiss, A. Mugge, E. Hofsteede-Botden, I. de Braak-Drent, and A. van der Kooij for their participation in the perceptual evaluation experiment.

REFERENCES

- Bokhout B, Hofman FXWM, van Limbeek J, Kramer GJC, Prah-Andersen B. Increased caries prevalence in 2.5-year-old children with cleft lip and/or palate. *Eur J Oral Sci.* 1996;104:518–522.
- Bzoch KR. Measurement and assessment of categorical aspects of cleft palate speech. In: Bzoch KR, ed. *Communicative Disorders Related to Cleft Lip and Palate*. Boston: Little, Brown; 1979.
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic Press; 1977.
- Dalston RM, Marsh JL, Vig KW, Witzel MA, Blumsted RM. Minimal standards for reporting the results of surgery on patients with cleft lip and palate, or both: a proposal. *Cleft Palate J.* 1988;25:3–7.
- Estrem T, Broen PA. Early speech production of children with cleft palate. *J Speech Hear Res.* 1989;32:12–23.
- Eurocleft Speech Group: Bronsted K, Grunwell P, Henningson G, Jansonius K, Karling J, Meijer M, Ordning U, Sell D, Wyatt R, Vermeij-Zieverink E. Cleft palate speech in a European perspective: Eurocleft Speech Project. In: Grunwell P, ed. *Analysing Cleft Palate Speech*. London: Whurr Publishers; 1993: 142–160.
- Folkins JW, Moon JB. Approaches to the study of speech production. In: Bardach J, Morris HL, eds. *Multidisciplinary Management of Cleft Lip and Palate*. Philadelphia: WB Saunders; 1990:707–716.
- Gnoinski W. Infant orthopedics and later orthodontic monitoring for unilateral cleft lip and palate patients in Zurich. In: Bardach J, Morris HL, eds. *Multidisciplinary Management of Cleft Lip and Palate*. Philadelphia: WB Saunders; 1990:578–585.
- Grunwell P, Russell J. Vocalisations before and after cleft palate surgery: a pilot study. *Br J Disord Commun.* 1987;22:1–17.
- Harding A, Grunwell P. Characteristics of cleft palate speech. *Eur J Disord Commun.* 1996;31:331–357.
- Hirschberg J, Van Demark DR. A proposal for standardization of speech and hearing evaluations to assess velopharyngeal function. *Folia Phoniatr Logop.* 1997;49:158–167.
- Hotz M, Gnoinski W, Perko M, Nussbaumer H, Hof E, Haubensak R. The Zurich approach, 1964 to 1984. In: Hotz M, Gnoinski W, Perko M, Nussbaumer H, Hof E, Haubensak R, eds. *Early Treatment of Cleft Lip and Palate*. Toronto: Hans Huber Publishers; 1986:42–48.
- Huddart AG. Presurgical orthopedic treatment in unilateral cleft lip and palate. In: Bardach J, Morris HL, eds. *Multidisciplinary Management of Cleft Lip and Palate*. Philadelphia: WB Saunders; 1990:574–578.
- Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care.* 1989;27:178–189.
- Kent RD. Hearing and believing: some limits to the auditory-perceptual assessment of speech and voice disorders. *Am J Speech Lang Pathol.* 1996; 5:7–23.
- Konst EM, Rietveld T, Peters H, Kuijpers-Jagtman AM. Language skills of young children with unilateral cleft lip and palate following infant orthopedics: a randomized clinical trial. *Cleft Palate Craniofac J.* 2003a;40:356–362.
- Konst EM, Rietveld T, Peters H, Prah-Andersen B. Phonological development of toddlers with unilateral cleft lip and palate who were treated with and without infant orthopedics: a randomized clinical trial. *Cleft Palate Craniofac J.* 2003b;40:32–39.
- Konst EM, Weersink-Braks H, Rietveld T, Peters H. An intelligibility assessment of toddlers with cleft lip and palate who received and did not receive presurgical infant orthopedic treatment. *J Commun Disord.* 2000;33:483–501.
- Konst EM, Weersink-Braks H, Rietveld T, Peters HFM. Prelexical development of unilateral cleft lip and palate babies with reference to presurgical infant orthopaedics: a randomized prospective clinical trial. *Clin Linguist Phon.* 1999;13:395–407.
- Kreiman J, Gerratt BR, Kempster GB, Erman A, Berke GS. Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res.* 1993;36:21–40.
- Lubit EC. Cleft palate orthopedics: why, when, how. *Am J Orthod.* 1976;69: 562–571.
- McNeil CK. Congenital oral deformities. *Br Dent J.* 1956;18:191–198.
- McWilliams BJ, Morris HL, Shelton RL. *Cleft Palate Speech*. Philadelphia: BC Decker Inc.; 1984.
- Morley ME. *Cleft Palate and Speech*. Edinburgh: Livingstone; 1970.
- O’Gara MM, Logemann JA. Phonetic analyses of the speech development of babies with cleft palate. *Cleft Palate J.* 1988;25:122–134.
- Oliver HT. Neonatal orthodontics. *Trans Eur Orthod Soc.* 1973;562–563.
- Prah C, Kuijpers-Jagtman AM, Van’t Hof MA, Prah-Andersen B. A randomised prospective clinical trial into the effects of infant orthopaedics on maxillary arch dimensions in unilateral cleft lip and palate (Dutchcleft). *Eur J Oral Sci.* 2001;109:297–305.
- Preminger JE, Van Tasell DJ. Quantifying the relation between speech quality and speech intelligibility. *J Speech Hear Res.* 1995;38:714–725.
- Pruzansky S. Pre-surgical orthopedics and bone grafting for infants with cleft lip and palate: A dissent. *Cleft Palate J.* 1964;1:164–186.
- Rietveld T, van Hout R. *Statistical Techniques for the Study of Language and Language Behaviour*. Berlin/New York: Mouton de Gruyter; 1993.
- Schiavetti N. Scaling procedures for the measurement of speech intelligibility. In: Kent RD, ed. *Intelligibility in Speech Disorders*. Amsterdam/Philadelphia: John Benjamins Publishing Co.; 1992:11–34.
- Schlichting JEPT, van Eldik MCM, Lutje Spelberg HC, van der Meulen SJ, van der Meulen BF. *Schlichting test voor Taalproductie*. Nijmegen: Berkhout Nijmegen BV; 1995.
- Sell D, Grunwell P, Mildinhal S, Murphy T, Cornish TAO, Bearn D, Shaw WC, Murray JJ, Williams AC, Sandy JR. Cleft lip and palate care in the United Kingdom—the clinical standards advisory group (CSAG) study. Part 3: speech outcomes. *Cleft Palate Craniofac J.* 2001;38:30–37.
- Sell D, Harding A, Grunwell P. A screening assessment of cleft palate speech (Great Ormond Street Speech Assessment). *Eur J Disord Commun.* 1994; 29:1–16.

