

B. Vogt<sup>1</sup> · A. Maier<sup>1</sup> · A. Batliner<sup>2</sup> · E. Nöth<sup>2</sup> · E. Nkenke<sup>3</sup> · U. Eysholdt<sup>1</sup> · M. Schuster<sup>1</sup>  
<sup>1</sup> Abteilung für Phoniatrie und Pädaudiologie, Universitätsklinikum Erlangen  
<sup>2</sup> Lehrstuhl für Mustererkennung, Technische Fakultät der Universität Erlangen-Nürnberg  
<sup>3</sup> Klinik und Poliklinik für Mund-, Kiefer-, Gesichtschirurgie, Universitätsklinikum Erlangen

# Numerische Quantifizierung der Verständlichkeit von Schulkindern mit isolierter und kombinierter Gaumenspalte

**Spaltfehlbildungen der Lippe, des Kiefers und des Gaumens zählen zu den häufigsten angeborenen orofazialen Fehlbildungen mit einer Prävalenz von 1,0–2,2/1000 Neugeborene [3]. Die strukturellen Veränderungen der Nase, des Rachens und des Mundes einschließlich des Kiefers können Folgen haben auf die Atmung, die Ernährung, das Sprechen, das Hören, die psychosoziale Kompetenz und die schulische Laufbahn [9, 16, 17].**

Lautbildungsstörungen gehören zu den häufigen Folgen einer Spaltfehlbildung. Zu den typischen Lautbildungsstörungen bei Spaltfehlbildungen, die den Gaumen betreffen, zählen die veränderte nasale Luftführung [15], Rückverlagerung oder Vorverlagerung der Artikulation und veränderte Sprechspannung [8, 9]. Besonders die veränderte Luftführung bei der Lautbildung scheint für die Verständlichkeit des betroffenen Kindes von herausragender Bedeutung zu sein. Meist besteht diese in einer *vermehrten* nasalen Schallführung (*Hyperrhinolalie*), die sich als offenes Näseln manifestiert und durch einen unzureichenden Abschluss des Nasenrachens durch das Gaumensegel entsteht. Aber auch die Verlagerung der Ar-

tikulation, Zahnfehlstellungen bei Kieferpalten und eine eingeschränkte Lippenfunktion können die Verständlichkeit wesentlich beeinflussen [29].

Für die Messung der Verständlichkeit existierten bisher keine objektiven Methoden. Es liegen zwar halbstandardisierte Erhebungsinstrumente für Lautbildungsstörungen vor wie z. B. der PLAKSS (Test zur *psycho*linguistischen Analyse kindlicher Sprechstörungen) [4] und andere Bildtafeln, welche jedoch nicht die spalttypischen Lautbildungsstörungen systematisch erfassen. Für die besonderen Bedingungen bei Patienten mit Spaltfehlbildungen wurde für Europa daher im Jahr 2000 ein Minimalstandard für die Diagnostik veröffentlicht [21]. Die Auswertung der Lautbildungsstörung erfolgt jedoch auch hierbei auditiv und bedeutet einen hohen zeitlichen und personellen Aufwand. Eine wesentliche Einflussgröße auf die Qualität der Bewertung ist bei diesem Vorgehen die Erfahrung des Bewerter [13], wobei selbst bei Erfahrenen noch eine große Intra-Rater-Varianz vorliegen kann [10].

Für wissenschaftliche Zwecke besteht der „objektive“ Goldstandard darin, die Störung durch mehrere Experten einstufen zu lassen, um mit dem daraus abgeleiteten Mittelwert oder Median ein in-

tersubjektiv abgesichertes auditives Urteil zu erhalten. Im klinischen Alltag erfolgt die Bewertung wegen des hohen Aufwands allerdings nicht durch mehrere Experten. Da die angewendeten Verfahren weiterhin stark variieren, ist auch die Vergleichbarkeit der Ergebnisse eingeschränkt [11]. Dieses klinisch gebräuchliche Vorgehen bedingt, dass die funktionellen Ergebnisse bei Kindern mit orofazialen Spaltfehlbildungen zwischen verschiedenen Zentren allein aus methodischen Gründen variieren können und wissenschaftlichen Kriterien wie auch den Forderungen nach „evidence based medicine“ nicht genügen.

Objektive oder apparative Verfahren, die schon länger verfügbar sind, erfassen lediglich Teilaspekte einer Kommunikationsstörung bei Spaltfehlbildung, nämlich die nasale Luftströmung bei der Phonation oder die Stimmqualität. Für die globale Bewertung der Sprechqualität existiert nun jedoch ein Verfahren, das die Verständlichkeit als Maß für die Sprechqualität objektiv quantifiziert [7, 18, 19]. Die Bewertung erfolgt hierbei untersucherunabhängig und gibt einen Wert auf einer linearen Skala an, im Gegensatz zur sonst üblichen kategorialen Bewertung auf einer mehrstufigen Skala oder quasilinearen visuellen Analogskala. Dieses neue

Tab. 1 Patienten und Kontrollgruppe

	Anzahl	Alter (Mittelwert, Standardabweichung)	Alter minimal	Alter maximal
Kontrollgruppe Schulalter	84	8;6±1;3	6;3	10;7
LKG unilateral	33	9;0±2;0	6;2	12;10
LKG bilateral	9	9;9±2;2	6;7	12;10
Gaumenspalte	10	8;7±2;3	6;0	12;3
Submuköse Gaumenspalte	6	8;7±1;7	6;2	10;3

Verfahren basiert auf der Spracherkennungstechnik.

Spracherkennungssysteme sind mittlerweile weit verbreitet (z. B. Telephonauskunftssysteme) und werden üblicherweise verwendet, um unbekannte Wörter oder Wortfolgen zu erkennen. Die Technik basiert auf statistischen Modellen, die die phonematischen und syntaktischen Besonderheiten der Sprache nutzen, z. B. welche Wörter am ehesten aufeinander folgen können.

Für unseren diagnostischen Zweck wird das System nun umgekehrt genutzt: aus einer bekannten Wortfolge von Einzelwörtern wird nun die Qualität des Gesprochenen bewertet, in dem das System prozentual angibt, wie viele der Wörter einer gesprochenen Wortfolge richtig erkannt wurden. Die Messung auf einer linearen Skala wird angegeben als Wortakkuratheit, dem Prozentsatz richtig erkannter Wörter, hier im Weiteren auch Verständlichkeitsgrad genannt. Eine Vorstudie ergab eine sehr hohe Übereinstimmung der apparativ quantifizierten Wortakkuratheit und der subjektiv durch mehrere Experten bestimmten Verständlichkeit bei Kindern mit Spaltfehlbildung ( $r=-0,9$ ,  $p<0,01$ ) [19].

In dieser prospektiven Studie werden nun erstmals objektive Daten mittels automatischer Spracherkennungstechnik zur Verständlichkeit von Schulkindern unter Berücksichtigung verschiedener Formen einer Spaltfehlbildung dargestellt und mit einer altersangepassten Kontrollgruppe verglichen. Durch den Vergleich des Verständlichkeitsgrads von Kindern mit isolierter Gaumenspalte und solchen mit kombinierten Spaltfehlbildungen werden die funktionellen Folgen in Bezug zum Ausmaß einer Spaltfehlbildung gesetzt.

## Methoden

### Patienten

Im Rahmen der interdisziplinären Sprechstunde am Spaltenzentrum der Universität Erlangen-Nürnberg wurden digitalisierte Sprachaufnahmen von 58 Kindern im Alter von 6;0–12;1 Jahren (Mittelwert 9;0±2;0 Jahre) unter Zustimmung der Eltern erstellt. Neun Kinder hatte eine doppelseitige, 33 Kinder eine einseitige Lippen-Kiefer-Gaumen-Spalte (LKG-Spalte), 10 Kinder eine Gaumenspalte, 6 Kinder eine submuköse Gaumenspalte (■ Tab. 1). Die Kinder waren alle deutschsprachig aufgewachsen, einige sprachen Dialekt.

Unter den Patienten mit einer isolierten Gaumenspalte war ein Kind mit einer Pierre-Robin-Sequenz, dessen kognitive Entwicklung sowie der Wortschatz überprüft und als altersentsprechend eingestuft wurden. Andere syndromale Erkrankungen waren bei den Kindern nicht bekannt. Die chirurgische Therapie aller Kinder orientierte sich am Vorgehen des Erlanger Spaltenzentrums: der Verschluss der Lippe erfolgt im Alter von 5–6 Monaten (Tennison-Randall), der des Gaumens wird im Alter von 10–12 Monaten (von Langenbeck u. Widmaier) und spätere osteoplastische Eingriffe werden zwischen dem 8. und 12. Lebensjahr durchgeführt.

### Kontrollgruppe

Die Kontrollgruppe bestand aus 84 Schulkindern einer Regelschule im Alter von 6;3–10;7 Jahren (8;6±1;3 Jahre) mit altersentsprechender Lautbildung. Entwicklungstypische Lautbildungsstörungen, z. B. während des Zahnwechsels, wurden als physiologisch angesehen und miteinbezogen. Die Kinder waren deutschsprachig, teils mit Dialekt, aufgewachsen.

## Testmaterial

Für die Sprachaufnahmen wurde das einheitliche Wortinventar des PLAKSS benutzt [4]. Der Test besteht aus 99 Bildtafeln und enthält alle Phoneme der deutschen Sprache an verschiedenen Positionen (An-, In- und Auslaut). Die Aufzeichnung erfolgte digital mit einem dnt Call 4U Comfort-Headset (16 kHz/16 bit).

### Spracherkennungssystem

Zur objektiven Messung der funktionellen Einschränkung aufgrund von Sprechstörungen kam in dieser Studie ein Spracherkennungssystem zur Anwendung, das bereits an einem ähnlichen Kollektiv überprüft wurde [19]. Sprachdaten werden dabei zuerst in kleine zeitliche Einheiten unterteilt (10–20 ms), dann werden deren temporale und spektrale Merkmale analysiert und kategorisiert und mit vorliegenden Wortmodellen verglichen (■ Abb. 1). Die Abfolge dieser Merkmale wird dann mit bekannten Wortmodellen verglichen, woraus berechnet werden kann, mit welchem bekannten Wortmodell das zu erkennende Signal am ehesten übereinstimmt.

Ein Spracherkennungssystem wertet 2 unterschiedliche Informationskanäle aus: Das Wissen über das akustische Signal wird mit stochastischen Wortmodellen (Hidden-Markov-Modelle) modelliert, das Wissen über die Abfolge von Wörtern einer Sprache wird mit stochastischen Sprachmodellen repräsentiert. Für unseren Zweck wurde nur Ersteres einbezogen, um die akustischen Eigenschaften des Gesprochenen stärker zu bewerten. Unser Erkennungssystem arbeitet auf akustischer Ebene polyphonbasiert, d. h. die akustischen Merkmale eines Phonems werden mit koartikulatorischen Modulationen des Phonems dargestellt.

Ein Spracherkennungssystem hat normalerweise ein sogenanntes Bi- oder Trigramm-Sprachmodell, d. h. die Wahrscheinlichkeit für ein Wort hängt von dem akustischen Signal (modelliert mit den Polyphon-Lautmodellen des Wortes) und der Wahrscheinlichkeit ab, dass dieses Wort auf das Vorgängerwort (Bigramm) oder die letzten beiden Vorgängerwörter

B. Vogt · A. Maier · A. Batliner · E. Nöth · E. Nkenke · U. Eysholdt · M. Schuster

### Numerische Quantifizierung der Verständlichkeit von Schulkindern mit isolierter und kombinierter Gaumenspalte

#### Zusammenfassung

**Hintergrund.** Spaltfehlbildungen können trotz adäquater Behandlung funktionelle Beeinträchtigungen, z. B. Lautbildungsstörungen, verursachen. Diese variieren individuell stark. Typisch sind z. B. veränderte nasale Luftführung und verlagerte Artikulation, die zu einer verminderten Verständlichkeit führen. Ein Zusammenhang zwischen der Art der Spaltfehlbildung und der Verständlichkeit konnte bisher nur durch kategoriale, subjektive Bewertungen beschrieben werden. In dieser Studie wird die spaltabhängige Verständlichkeit erstmals objektiv und numerisch mittels automatischer Spracherkennungstechnik quantifiziert.

**Methode.** Die spaltabhängige Verständlichkeit wurde objektiv und numerisch mittels automatischer Spracherkennungstechnik an Sprachaufnahmen des PLAKSS (Test

zur psycholinguistischen Analyse kindlicher Sprechstörungen) von 58 Kindern im Alter von  $9\pm 2$  Jahren quantifiziert. 33 Kinder hatten eine unilaterale Lippen-Kiefer-Gaumenspalte, 9 eine bilaterale Lippen-Kiefer-Gaumenspalte, 10 eine Gaumenspalte und 6 eine submuköse Gaumenspalte. Als Vergleich dienen Aufnahmen von 84 Kindern mit physiologischer Lautbildung einer Regelschule im Alter von  $9\pm 1$  Jahren. Der Sprachverständlichkeitsgrad wird als Prozentsatz korrekter verstandener Wörter einer Wortkette, der Wortakkuratheit, angegeben.

**Ergebnisse.** Die Wortakkuratheiten betragen bei der Kontrollgruppe zwischen 39 und 84% (Mittelwert  $62\pm 10\%$ ) und zwischen 0 und 73% (Mittelwert  $42\pm 17\%$ ) bei Kindern mit Spaltfehlbildung. Die Verständlichkeit der Kinder mit isolierter Gaumenspalte und der-

jenigen mit kombinierter Spaltfehlbildung unterscheidet sich nicht. Es besteht jedoch ein Unterschied jeder Gruppe zur Kontrollgruppe ( $p < 0,01$ ).

**Fazit.** Bei Kindern mit Gaumenspalte haben zusätzliche Spaltbildungen von Kiefer und Lippe keinen signifikanten Einfluss auf den Sprachverständlichkeitsgrad. Dies wurde erstmals mit objektiven Methoden quantifiziert. Der Bedarf an Diagnostik und Therapie bei Sprechstörungen ist bei Kindern mit isolierter Spaltfehlbildung des Gaumens genauso hoch wie bei kombinierten Spaltfehlbildungen.

#### Schlüsselwörter

Lippen-Kiefer-Gaumenspalten · Isolierte Gaumenspalte · Automatische Spracherkennung · Verständlichkeit · Objektive Bewertung.

### Numeric quantification of intelligibility in schoolchildren with isolated and combined cleft palate

#### Abstract

**Objective.** Even after adequate surgical and nonsurgical treatment, combined or isolated clefts often cause functional disorders, such as speech disorders. Speech disorders vary widely in extent and can take the form, for example, of specific features of articulation, with altered nasal emission and shifted articulation, leading to reduced speech intelligibility. So far it has not been possible to describe the relationship between cleft type and intelligibility except subjective, categorical evaluation.

**Methods.** Intelligibility of the speech as influenced by cleft palate in 58 children aged  $9\pm 2$  was quantified objectively and numerically by means of automatic speech recognition technology in speech recordings for the PLAKSS test [test for psycholinguistic analy-

sis of childhood speech impairments]. It was found that 33 children had a unilateral cleft lip and palate (CLP), 9, bilateral CLP, 10, isolated cleft palate, and 6, submucosal cleft palate. As a control group, 84 children aged  $9\pm 1$  years and with physiological articulation were recruited from a mainstream school. Speech intelligibility is expressed as word accuracy, which means the percentage of correctly recognized words in a word sequence.

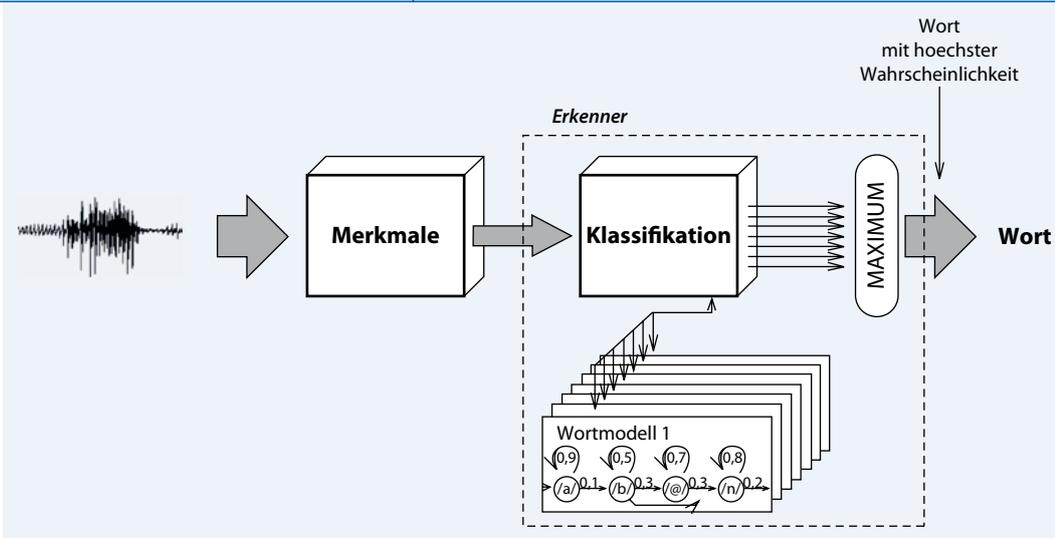
**Results.** Word accuracy levels in the control group were between 39% and 84% (mean  $62\pm 10\%$ ), while the children with clefts achieved values between 0% and 73% (mean  $42\pm 17\%$ ). There was no difference in intelligibility between children with isolated cleft palate and those with combined cleft type.

Differences were, however, found between each group and the control group ( $p < 0.01$ ).

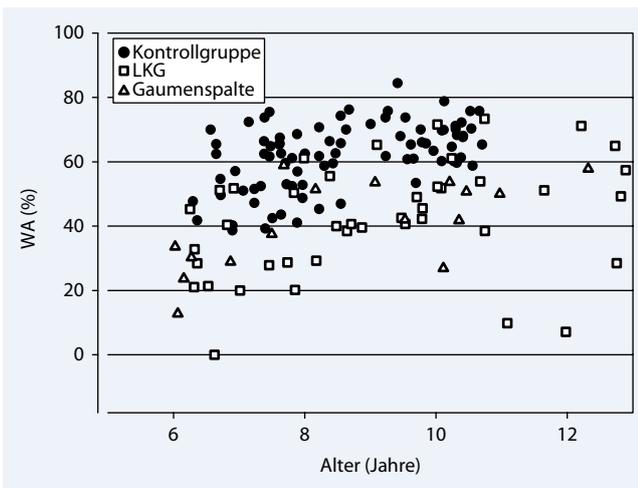
**Conclusion.** Speech intelligibility of children with isolated cleft palate does not significantly differ from that of children with combined cleft type including the palate, as has been objectively quantified for the first time by objective methods. In the presence of speech disorders, the need for diagnosis and therapy is just as great in children with isolated cleft palate as in children with combined cleft types.

#### Keywords

Cleft lip and palate · Isolated cleft palate · Automatic speech recognition · Speech intelligibility · Objective evaluation



**Abb. 1** ▲ Schematische Darstellung der Arbeitsweise eines herkömmlichen Spracherkennungssystems: Ablauf von links nach rechts (Erläuterung s. Text). Für die hier beschriebene diagnostische Methode erfolgt der Weg umgekehrt von rechts nach links: bei bekannten Wörtern lässt der Anteil korrekt erkannter Wörter Rückschlüsse auf die Qualität des akustischen Signals zu. (Nach [30])



**Abb. 2** ▲ Wortakkuraterie (WA) in Bezug zum Alter bei der Kontrollgruppe und den Kindern mit isolierter und kombinierter Gaumenspalte

(Trigramm) folgt. Wir benutzen jedoch lediglich ein sog. Unigramm-Sprachmodell (kontextfreie Auftrittswahrscheinlichkeit), um das Ergebnis jedes einzelnen Wortmodells zu gewichten [6], wodurch das akustische Signal stärker beachtet wird. Die Test-Perplexität (durchschnittliche Anzahl von Wörtern, die auf ein bestimmtes Wort innerhalb einer Sequenz folgen können) des Unigramm-Sprachmodells beträgt 94, was ausreichend ist, um eine schlechte Verständlichkeit als niedrige Worterkennungsrate (WA) abzubilden und umgekehrt. Eine weitere Verbesserung des Spracherkennungsergebnisses wurde durch sprechab-

hängige Adaption („maximum likelihood linear regression“, MLLR [5]) erreicht.

Das Sprachmodell wurde zuvor mit der Transliteration des gesprochenen Textes trainiert. Somit besaß das Spracherkennungssystem Informationen über die Häufigkeit des Auftretens jedes einzelnen Wortes im verwendeten Text. Die Trainingsdaten für das akustische Modell bestanden aus Sprachaufnahmen von 6,9 h Dauer von 23 Jungen und 30 Mädchen zwischen 10 und 14 Jahren einer lokalen Schule sowie aus vokaltraktnormierter Erwachsenensprache [23, 26]. Die dialektale Färbung entsprach der der Patienten und der Kontrollgruppe.

Die Leistungsfähigkeit eines Spracherkennungssystems wird üblicherweise mit der Wortakkuraterie (WA) gemessen. Die WA gibt an, wie stark eine erkannte Wortsequenz mit der gesprochenen Äußerung übereinstimmt, d. h. sie stellt den Prozentsatz an korrekt erkannten Wörtern dar. Diese Zahl wird durch die Anzahl der durch ein Spracherkennungssystem fälschlich eingefügten Wörter korrigiert.

Sie errechnet sich durch folgende Formel:

$$WA[\%] = 100 * (NC - NW) / N$$

NC: Anzahl der korrekt erkannten Wörter

NW: Anzahl der durch das Spracherkennungssystem falsch eingefügten Wörter

N: Anzahl aller gesprochenen Wörter

In der vorgestellten Anwendung sind sowohl das Spracherkennungssystem als auch das Sprachmaterial gleich. Damit ist die Variation der WA lediglich auf unterschiedliche Güte des Gesprochenen zurückzuführen. Mithilfe des Spracherkennungssystems errechneten wir die WA der Kinder aus den aufgenommenen Sprachdaten des PLAKSS-Tests; die Sprachdaten wurden zuvor transliteriert.

## Analyse und Auswertung der Daten

Die statistische Analyse erfolgte mittels Kolmogorov-Smirnov-Test, t-Test mit unabhängigen Stichproben sowie ANOVA (Software SPSS und SYSTAT). Für die Bestimmung der pathologischen Sprachverständlichkeit wurde ein „hartes“ Kriterium entsprechend anderen Tests zur Sprachentwicklung gewählt: eine Abweichung von mehr als 2 Standardabweichungen vom Mittelwert der Kontrollgruppe bei den altersbereinigten Daten gilt in dieser Studie als auffällig verminderte Verständlichkeit.

## Ergebnisse

Die Wortakkuratheit in den Aufnahmen variiert stark (■ **Tab. 2**). Im Normkollektiv beträgt die WA (nicht altersbereinigt) zwischen 39 und 84% (Mittelwert:  $62\% \pm 10\%$ ). Die größte WA in der Gruppe der Spaltpatienten ergibt sich mit 73%, die niedrigste mit 0% (Mittelwert:  $42\% \pm 17\%$ ). Der Mittelwert für jeden Spalttyp ist in ■ **Tab. 2** ersichtlich.

Die Wortakkuratheit wird mit zunehmendem Alter in der Kontrollgruppe signifikant größer (Steigung = 3,8,  $p < 0,01$ ), was sich bei den Kindern mit Spaltfehlbildung nicht nachweisen lässt (■ **Abb. 2**). Für die weitere Berechnung der Unterschiede zwischen den verschiedenen Typen der Spaltfehlbildung wurden die Wortakkuratheiten entsprechend der Kontrollgruppe mit der folgenden Formel altersbereinigt:  $WA_{\text{altersbereinigt}} = WA - (\text{Alter} \times 3,8)$ , wobei der Wert 3,8 hier der Steigung bei der Korrelation WA vs. Alter in der Kontrollgruppe entspricht (■ **Abb. 2**).

Die Wortakkuratheit bei Kindern mit einseitiger oder beidseitiger LKG-Spalte ( $p=0,93$ ) weicht nicht signifikant von der bei Kindern mit submuköser oder offener Gaumenspalte ab ( $p=0,9$ ). Betrachtet man die Patienten mit ein- oder beidseitiger LKG-Spalte zusammengefasst in einer Gruppe im Vergleich zu den Kindern mit submuköser oder offener Gaumenspalte ebenfalls in einer gemeinsamen Gruppe, so bestehen auch hier keine signifikant unterschiedlichen WA ( $p=0,78$ ; ■ **Abb. 3**).

**Tab. 2** Ergebnisse der Sprachverständlichkeitsprüfung

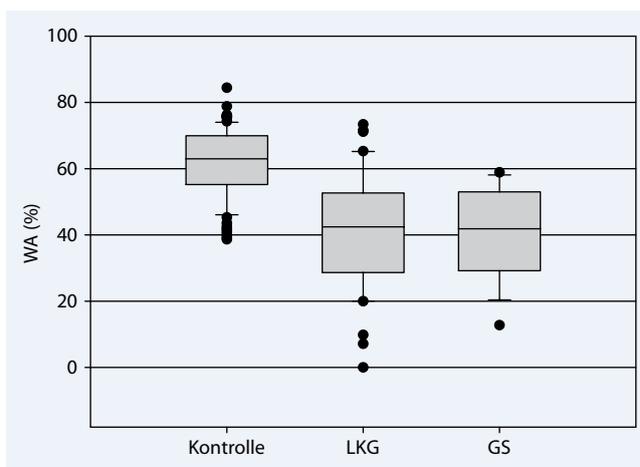
	Anzahl	WA (M, SD) [%]; Max., Min. [%]	WA altersbereinigt (M, SD) [%]
Kontrollgruppe Schulalter	84	62±10 84; 39	29±9
LKG unilateral	33	42±16 72; 7	8±16
LKG bilateral	9	42±24 73; 0	4±19
Gaumenspalte	10	41±16 59; 13	8±10
Submuköse Gaumenspalte	6	40±11 51; 27	7±12

WA Wortakkuratheit, M Mittelwert, SD Standardabweichung, Max. maximaler Wert, Min. minimaler Wert.

**Tab. 3** Häufigkeit der pathologisch verminderten Verständlichkeit

	n	%	%
LKG unilateral	20/33	61	60
LKG bilateral	5/9	56	
Gaumenspalte	5/10	50	56
Gaumenspalte submukös	4/6	67	

LKG Lippen-Kiefer-Gaumen-Spalte.



**Abb. 3** ▲ Boxplot der Wortakkuratheit bei Schulkindern der Kontrollgruppe (84 Kinder), mit Lippen-Kiefer-Gaumen-Spalten (LKG, 42 Kinder) und isolierter Gaumenspalte (GS, 16 Kinder). Der Mittelwert ist in jeder Box durch einen Querstrich angegeben. Die Box beinhaltet die Kinder innerhalb der mittleren Quartile. Die Punkte geben jeden Wert außerhalb der Standardabweichungen jeder Gruppe an

■ **Tab. 3** gibt die Anzahl der Kinder an, bei denen die WA mindestens 2 Standardabweichungen geringer als bei der Kontrollgruppe ist.

## Diskussion

Die Ermittlung der Verständlichkeit dient als globale Bewertung der Sprechqualität, also der Qualität der phonematischen Ausformung von Wörtern. In dieser Studie quantifizieren wir die Verständlichkeit bei verschiedenen Spalttypen bei Kindern

im Schulalter erstmals mit einer objektiven und apparativen Methode, der automatischen Spracherkennungstechnik.

Automatische Spracherkennung ist von 5 Faktoren abhängig: vom Sprecher, vom Sprechregister (Sprache beim Lesen, Spontansprache), vom Vokabular, von der grammatischen Komplexität oder Perplexität (durchschnittliche Anzahl von Wörtern, die auf ein bestimmtes Wort innerhalb einer Sequenz folgen können) und von den akustischen Eigenschaften des Aufnahmegeäts [6]. Der Einfluss der

meisten dieser Faktoren kann durch Anwendung eines Standard-Tests – wie des PLAKSS-Tests – und durch stabile Aufnahmebedingungen – wie in dieser Studie zur diagnostischen Anwendung – minimiert werden. Demnach bleibt der Sprecher als hauptsächlicher Faktor bestehen. Dabei dient der Prozentsatz korrekt erkannter Wörter einer Wortfolge (Wortakkuratheit) als Maß für den Verständlichkeitsgrad.

Natürlich ist die Wortakkuratheit eigentlich nicht gleichbedeutend mit der Verständlichkeit. Beide werden zwar von der Stimmqualität, der phonetischen und morphosyntaktischen Struktur, von Hintergrundgeräuschen, Lautstärke und Sprechgeschwindigkeit beeinflusst. Die Verständlichkeit umfasst jedoch auch den „menschlichen Faktor“.

Selbst wenn man nicht jedes einzelne Wort oder jede Silbe einer gesprochenen Sequenz versteht, kann die Bedeutung mithilfe des Kontexts und pragmatischer und prosodischer Elemente der Sprache erschlossen werden. Naturgemäß ist die perzeptive Beurteilung durch unterschiedliche kognitive Voraussetzungen, unterschiedliche Sprachkompetenz und Erfahrung der Untersucher uneinheitlich und nur bedingt verlässlich, wie dies auch von Paal et al. und Keuning et al. für diese Patientengruppe beschrieben wurde [10, 13]. Preminger und Van Tasell beschreiben zwar eine hohe Korrelation zwischen verschiedenen Bewertern, zeigen aber auch, dass bei diskreter Skala keine hohe Übereinstimmung bei der Wahl der Kategorien besteht, was sich mit den üblichen statistischen Verfahren wie der Korrelationsanalyse nicht gut erfassen lässt [14]. Daher sind zur Beurteilung subjektiver Beurteilungen durch mehrere Bewerter weitere Kennzahlen für die Prüfung der Reliabilität erforderlich wie Multi-Rater- $\kappa$  oder Cronbach- $\alpha$ .

Die automatische Spracherkennung hingegen basiert auf der statistischen Modellierung von akustisch-phonetischem und linguistischem Wissen, wobei für diese Anwendung der Schwerpunkt auf den akustik-phonetischen Hintergrund gelegt wurde. Die Berechnung der Wortakkuratheit vermeidet durch Wertung jedes einzelnen Wortes eine unsystematische Verfälschung einer subjektiven Bewertung

und stellt sich in einer Vorarbeit [19] als übereinstimmend dar zur perzeptiven Beurteilung bei Einzelwörtern. Da sie unabhängig misst, weichen die Ergebnisse einer Wiederholungsmessung am selben Sprachmaterial nicht voneinander ab. Sie kann daher als genaues objektives Maß für den Sprachverständlichkeitsgrad gelten.

Im Vergleich zu anderen Studien [24], die sich mit Lautbildungsstörungen oder der Verständlichkeit von Kindern mit Spaltfehlbildungen auseinandersetzen, wurde in dieser Studie eine objektive Methode im Vergleich zu einer Kontrollgruppe gleichen Alters angewandt. Es wurden Kinder in einem Alter gewählt, in dem prinzipiell das komplette Lautinventar korrekt artikuliert wird, abgesehen von physiologischen Beeinträchtigungen wie dem Zahnwechsel, und in dem die primären chirurgischen Maßnahmen abgeschlossen sind, aber sich die Frage nach sprechverbessernden Operationen stellen kann. Entsprechend den Angaben zur Inzidenz in der Literatur finden sich auch in dieser Studie mehr Kinder mit kombinierter Spaltfehlbildung als mit isolierter Gaumenspalte [12, 22]. Die Gruppengröße ist verglichen mit den genannten Studien angemessen und lässt besonders bei klarer Fragestellung und großem Vergleichskollektiv eine eindeutige Aussage zu, belegt am Signifikanzniveau.

Diese Studie konzentriert sich ausschließlich auf die Verständlichkeit und berücksichtigt damit das Spektrum der verschiedenen Lautfehlbildungen nicht, auch wenn dies in einigen Publikationen gefordert wird [2, 20, 28]. Allerdings ist die Verständlichkeit der wesentliche Ergebnisparameter für die Lautbildung, an dem sich die Therapie von Spaltfehlbildungen messen lassen muss. Diese konnte hier erstmals mit einem objektiven Maß quantifiziert werden.

Zur Bedeutung verschiedener spezieller Lautbildungsstörungen für die Verständlichkeit liegen bisher unterschiedliche Ergebnisse vor [27]. Hierzu ist jedoch der Einfluss der jeweils gesprochenen Sprache gerade beim Vergleich mit Kollektiven mit anderer Muttersprache nicht bestimmbar und die statistische Aufarbeitung durch eine subjektive, kategoriale Bewertung eingeschränkt. Im Gegensatz zu der früher üblichen subjektiven

Bewertung bieten sich mit der hier angewandten objektiven Methode der Verständlichkeitsbewertung weitaus mehr Möglichkeiten statistischer Vergleiche, da die Ungenauigkeiten einer nichtlinearen Skalierung entfallen. Das Objektivieren des Verständlichkeitsgrads ist der 1. Schritt, mit dem die Spracherkennungstechnik die Diagnostik verbessern kann. Im Rahmen eines von der Deutschen Forschungsgemeinschaft geförderten Projekts werden wir weitere Variablen einbeziehen, um dann den Zusammenhang zwischen verschiedenen Formen und Gradierungen von Lautbildungsstörungen und der Verständlichkeit eindeutig belegen zu können.

Bei der Konzeption dieser Studie wurde entsprechend früheren Studien zu Sprechstörungen bei Spaltfehlbildungen eine große Variabilität der Verständlichkeit angenommen. Tatsächlich zeigte sich dies auch in unserer Untersuchung unter Beachtung spalttyp- und altersabhängiger Faktoren.

Aus pathomorphologischer Sicht wäre eine schlechtere Wortakkuratheit bei Kindern mit LKG-Spalte im Vergleich zu Kindern mit einer isolierten Gaumenspalte anzunehmen, da zusätzlich zum Gaumendefekt noch eine Störung im Bereich des Kiefers und der Lippe vorliegt, die einen Einfluss auf die Lautbildung in den vorderen Artikulationszonen oder eine Verlagerung der Artikulation bewirken könnte. Unsere Ergebnisse konnten jedoch keinen signifikanten Unterschied zwischen Kindern mit LKG-Spalte und Kindern mit einer isolierten Gaumenspalte nachweisen. Eine Fehlbildung im Bereich der vorderen Artikulationszone scheint also nicht bedeutsam für die Verständlichkeit der Kinder zu sein, wofür auch die Ergebnisse einer anderen Studie sprechen [24], obwohl eine Fehlbildung der Lippen und des Alveolarkammes einen negativen Einfluss auf die Artikulation haben können. Der Zusammenhang zwischen Lautfehlbildungen der vorderen Artikulationszone und der Verständlichkeit ist aber möglicherweise inkonsistent und bislang nicht ausreichend dargelegt worden. Dies wird im Weiteren nun gesondert in einer größeren Studie gezielt untersucht werden.

Auch zwischen unilateraler und bilateraler LKG-Spalte ließen sich in dieser Stu-

die keine signifikanten Unterschiede feststellen, was durchaus den Angaben der Arbeitsgruppe von Van Lierde entspricht [25]. Allerdings fanden Dalston et al. [2] bei Kindern mit beidseitiger LKG-Spalte vermehrte Lautfehlbildungen durch kompensatorische Artikulation, wodurch sich schlussfolgernd negative Auswirkungen auf die Verständlichkeit in dieser Patientengruppe erwarten ließen. Eine Überprüfung dieser These ist an einem größeren Studienkollektiv geplant.

Die Beschreibung einer Kontrollgruppe gleicher Altersstruktur ermöglicht die Festlegung eines Normbereichs der Verständlichkeit mit dieser neuen Methode und die Erfassung derer, die außerhalb der Norm liegen. Mit der kategorialen Bewertung konnte dies bei der subjektiven Bewertung bisher umgangen werden; allerdings konnte die Abweichung damit dann auch nur unscharf angegeben werden.

Sell [20] stellte in einer Patientengruppe von 12-Jährigen mit unilateraler LKG-Spalte fest, dass 47% der Patienten alterssprechend verständlich waren, während das Sprechen von 19% der Kinder entweder unmöglich zu verstehen, gerade noch verständlich für Fremde oder so verschieden von dem anderer Kinder war, dass dadurch kritische Kommentare provoziert wurden. In einer Studie von Timmons [24] wurden Kinder mit LKG-Spalte und mit isolierter Gaumenspalte im Alter von 5–12 Jahren untersucht. 53% der Patienten mit LKG-Spalte und 37% der Gaumenspaltpatienten waren normal verständlich, während das Sprechen von 12% der Gruppe mit LKG-Spalte und 30% der Patienten mit Gaumenspalte gerade noch verständlich bzw. so verändert war, dass kritische Bemerkungen folgten. In beiden Studien ist die Verständlichkeit von wenigstens etwa der Hälfte aller untersuchten Spaltpatienten abweichend von der Norm.

In unserer Untersuchung ist bei 56–60% der Kinder der Verständlichkeitsgrad, angegeben als Wortakkuratheit, auffällig vermindert im Vergleich zur Kontrollgruppe (■ **Tab. 3**). Man kann diese Werte allerdings nicht uneingeschränkt mit den oben genannten, subjektiven kategorialen Bewertungen der anderen Arbeitsgruppen vergleichen, da in der vorliegenden Studie eine lineare Skala sowohl

bei den Patienten als auch bei der Kontrollgruppe zur Anwendung kam. Hierbei verwendeten wir zur Definition einer auffälligen Verständlichkeit ein bei Sprachtests übliches Kriterium, nämlich eine Abweichung von mehr als 2 Standardabweichungen vom Mittel der Norm.

Besonders bei Kindern und Jugendlichen ist die Wortakkuratheit nicht selbstverständlich gleich der von Erwachsenen. Es besteht eine Altersabhängigkeit der Verständlichkeit sowie der Wortakkuratheit, gerade im Verlauf der Sprachentwicklung und der physiologischen Veränderung der Lautbildung, was beim hier vorliegenden Studiendesign berücksichtigt werden konnte. Somit können Ungenauigkeiten durch die sonst angewandte kategoriale Bewertung, bei der kein Bezug zum Alter hergestellt wird, ausgeschlossen werden, sowie Zusammenhänge deutlicher gemacht werden [1]. Für die Festlegung eines allgemeinen altersabhängigen Standards ist für die künftigen Studien eine Vergrößerung des Kontrollkollektivs vorgesehen.

Bei der kritischen Durchsicht der Ergebnisse besonders der älteren Kinder (■ **Abb. 2**) kommt jedoch der Verdacht auf eine Verfälschung dieser Ergebnisse auf, da viele eine niedrige WA haben. Eine Erklärung mag in der Struktur der Behandlung liegen: Bei den untersuchten Kindern handelt es sich um die Patienten einer hochspezialisierten Sprechstunde, die einen großen Einzugsbereich umfasst. Für viele Kinder bedeutet dies einen langen Anfahrtsweg. Patienten, bei denen keine aktuellen Behandlungen geplant sind und keine Probleme vorliegen, kommen aus diesem Grund möglicherweise nicht mehr zur regelmäßigen Untersuchung, sodass der Anteil der Kinder, die eine Sprechstörung haben, im Kollektiv der Spaltstunde überproportional hoch sein könnte. Dies könnte auch erklären, weshalb sich bei den Patienten keine Altersabhängigkeit der Wortakkuratheit zeigt. Die mittlerweile schon mögliche dezentrale Erfassung der Verständlichkeit am PC auch außerhalb der spezialisierten Sprechstunde könnte die Datenerfassung gerade für die weit entfernt lebenden Kinder erleichtern und so die regelmäßige Kontrolle der Sprachentwicklung auch dieser Kindern gewährleisten.

## Fazit

**Die Spracherkennungstechnik zur objektiven Bewertung der Verständlichkeit hat sich, wie auch schon in der Pilotstudie [19], als aussagekräftige, dabei zeitsparende diagnostische Methode bewährt. Es besteht kein Unterschied des Verständlichkeitsgrads zwischen Patienten mit isolierter Gaumenspalte und kombinierter Spaltfehlbildung mit Beteiligung des Gaumens. Aus der großen Anzahl auffällig vermindert verständlicher Kinder leitet sich ein hoher therapeutischer Bedarf ab. Die Indikation und Art der Therapie (konservative oder chirurgische Therapie) sollte für den Einzelfall innerhalb der spezialisierten Sprechstunde geklärt werden. Der eingeschränkte Verständlichkeitsgrad vieler Kinder mit isolierter Spaltfehlbildung des Gaumens sollte bei der Festlegung des Grades der Behinderung berücksichtigt werden. Die Festlegung des GdB kann für den Einzelfall mit objektiven Daten erleichtert werden.**

## Korrespondenzadresse

### M. Schuster

Abteilung für Phoniatrie und Pädaudiologie  
Universitätsklinikum Erlangen  
Bohlenplatz 21, 91054 Erlangen  
maria.schuster@phoni.imed.uni-erlangen.de

**Danksagung.** Diese Arbeit wurde von der Johannes- und Frieda-Marohn-Stiftung der Friedrich-Alexander-Universität Erlangen-Nürnberg sowie der Deutschen Forschungsgemeinschaft (SCHU 2320/1–1) unterstützt. Die Autoren sind für den Inhalt dieses Artikels verantwortlich.

**Interessenkonflikt.** Der korrespondierende Autor gibt an, dass kein Interessenkonflikt besteht.

## Literatur

1. Cummins RA, Gullone E (2000) Why we should not use 5-point Likert scales: The case for subjective quality of life measurement. Proceedings, Second International Conference on Quality of Life in Cities. National University of Singapore, Singapore, pp 74–93
2. Dalston RM, Warren DW, Dalston ET (1992) A preliminary study of nasal airway patency and its potential effect on speech performance. Cleft Palate Craniofac J 29: 330–335
3. Derijcke A, Eerens A, Carels C (1996) The incidence of oral cleft: a review. Br J Oral Maxillofac Surg 34: 488–494
4. Fox AV (2002) PLAKSS – Psycholinguistische Analyse kindlicher Sprechstörungen. Swets & Zeitlinger, Frankfurt am Main

5. Gales MFJ, Pye D, Woodland PC (1996) Variance compensation within the MLLR framework for robust speech recognition and speaker adaptation. In: Proc ICSLP, 96, vol 3. Philadelphia, USA, pp 1832–1835
6. Gallwitz F, Niemann H, Nöth E (1999) Spracherkennung – Stand der Technik, Einsatzmöglichkeiten und Perspektiven. *Wirtschaftsinformatik* 41: 538–547
7. Haderlein T, Wittenberg T, Decher M, Nöth E (2001) Automatische Stotterererkennung und Klassifikation mit Hilfe von Hidden-Markov-Modellen. In: Gross M, Kruse E (Hrsg) Aktuelle phoniatriisch-pädaudiologische Aspekte 2000/2001. Median, Heidelberg, S 124–130
8. Harding A, Grunwell P (1998) Active versus passive cleft-type speech characteristics. *Int J Lang Commun Disord* 33: 329–352
9. Harding A, Grunwell P (1996) Characteristics of cleft palate speech. *Eur J Disord Commun* 31: 331–357
10. Keuning K, Wieneke G, Dejonckere P (1999) The intrajudge reliability of the perceptual rating of cleft palate speech before and after pharyngeal flap surgery: The effect of judges and speech samples. *Cleft Palate Craniofac J* 36: 328–333
11. Lohmander A, Olsson M (2004) Methodology for perceptual assessment of speech in patients with cleft palate: a critical review of the literature. *Cleft Palat Craniofacial J* 41: 64–70
12. Owens JR, Jones JW, Harris F (1985) Epidemiology of facial clefting. *Arch Dis Child* 60: 521–524
13. Paal S, Reulbach U, Strobel-Schwarthoff K et al. (2005) Bewertung von Sprechaffälligkeiten bei Kindern mit LKG-Spalten. *J Orofac Orthop* 66: 270–278
14. Preminger JE, Van Tasell DJ (1995) Quantifying the relation between speech quality and speech intelligibility. *J Speech Hear Res* 38: 714–725
15. Pulkkinen J, Haapanen ML, Paaso M et al. (2001) Velopharyngeal function from the age of three to eight years in cleft palate patients. *Folia Phoniatri Logop* 53: 93–98
16. Rosanowski F, Eysholdt U (2002) Phoniatrie aspects in cleft lip patients. *Facial Plast Surg* 18: 197–203
17. Schönweiler R, Schönweiler B, Schmelzeisen R (1994) Hörvermögen und Sprachleistungen bei 417 Kindern mit Spaltfehlbildungen. *HNO* 42: 691–696
18. Schuster M, Haderlein T, Nöth E et al. (2006) Intelligibility of laryngectomees' substitute speech: Automatic speech recognition and subjective rating. *Eur Arch Otorhinolaryngol* 263: 188–193
19. Schuster M, Maier A, Haderlein T et al. (2006) Evaluation of speech intelligibility for children with cleft lip and palate by means of speech recognizer technique. *Int J Pediatr Otorhinolaryngol* 70: 1741–1747
20. Sell D, Grunwell P, Mildinhal S et al. (2001) Cleft lip and palate care in the United Kingdom – the Clinical Standards Advisory Group (CSAG) Study. Part 3: speech outcomes. *Cleft Palate Craniofac J* 38: 30–37
21. Shaw WC, Semb G, Nelson P et al. (2000) The Eurocleft Project 1996–2000. Standards of care for cleft lip and palate in Europe. European Commission Biochemical and Health Research. IOS Press, Amsterdam
22. Sipek A, Gregor V, Horacek J, Masatova D (2002) Facial clefts from 1961 to 2000 – incidence, prenatal diagnosis and prevalence by maternal age. *Ceska Gynekol* 67: 260–267
23. Stemmer G, Hacker C, Steidl S, Nöth E (2003) Acoustic normalization of children's speech. In: Proc. European Conference on Speech Communication and Technology, vol 2. Geneva, Switzerland, pp 1313–1316
24. Timmons MJ, Wyatt RA, Murphy T (2001) Speech after repair of isolated cleft palate and cleft lip and palate. *Br J Plast Surg* 54: 377–384
25. Van Lierde KM, De Bodt M, Van Borsel J et al. (2002) Effect of cleft type on overall speech intelligibility and resonance. *Folia Phoniatri Logop* 54: 158–168
26. Wahlster W (ed) (2000) *Verbmobil: Foundations of speech-to-speech translation*. Springer, Berlin
27. Whitehill TL, Chau CH (2004) Single-word intelligibility in speakers with repaired cleft palate. *Clin Linguist Phon* 18: 341–355
28. Witzel MA (1991) Speech evaluation and treatment. *Atlas Oral Maxillofac Surg Clin North Am* 3: 501–516
29. Wohlleben U (2004) Die Verständlichkeitsentwicklung von Kindern mit Lippen-Kiefer-Gaumen-Segel-Spalten. Schulz-Kirchner, Idstein
30. Schukat-Talamazzini EG (1995) *Automatische Spracherkennung – Grundlagen, statistische Modelle und effiziente Algorithmen*. Künstliche Intelligenz. Vieweg, Braunschweig