

COMPUTERGESTÜTZTE BESTIMMUNG DES SPRECHFLUSSES BEI VORSCHULKINDERN

Valentin Kany¹, Jürgen Trouvain²

Sprachwissenschaft und Sprachtechnologie, Universität des Saarlandes

¹*valentin.kany@uni-saarland.de*, ²*trouvain@lst.uni-saarland.de*

Kurzfassung: Die vorliegende Studie untersucht, inwiefern sich die in der Phonetik verbreiteten Skripte zur automatisierten Feststellung von verschiedenen Aspekten des Sprechflusses von de Jong et al. zur Beurteilung des Sprachstands bei Kindern eignen und in welcher Art und Weise die Methodik angepasst werden könnte. Dazu wurden Sprachdaten von Vorschulkindern mit Deutsch als Muttersprache bzw. als Zweitsprache mithilfe eines Serious Game zur Sprachstandserhebung elizitiert. Die Audiodaten wurden bezüglich Artikulationsrate, Pausen und Füllpartikeln sowohl automatisiert durch die Skripte als auch manuell annotiert. Die Ergebnisse zeigen, dass sich die Skripte zur Ermittlung der Artikulationsrate mit einer relativ hohen Übereinstimmung mit der manuellen Ermittlung zur Verwendung in Sprachstandserhebungsverfahren eignen. Auch die automatische Erkennung von Sprechpausen weist einen hohen Precision-Wert auf und könnte als Instrument in Sprachstandserhebungen verwendet werden. Eine solche Verwendung würde mit einer Erweiterung um die hier vorgestellte manuelle Methode zum Annotieren von disfluenten und nicht disfluenten Pausen profitieren. Bei den Füllpartikeln stellte sich die automatisierte Klassifizierung hingegen als weniger geeignet heraus. Hier wurde keine hohe Übereinstimmung mit der menschlichen Annotation festgestellt. Um in der Praxis Verwendung finden zu können, muss diese Methode noch erweitert werden, zum Beispiel durch Miteinbeziehen der Pausenerkennung.

1 Einleitung

Fließendes Sprechen stellt einen wesentlichen Indikator für Sprachkompetenz dar, weswegen der Sprechfluss in Sprachstandserhebungsverfahren einen relevanten Aspekt verkörpert [1]. Der Grad des Sprechflusses ist Schwankungen unterworfen, was sich auch in Unterbrechungen des Sprechflusses bemerkbar macht. Diese Disfluenzen können verschiedener Natur sein, beispielsweise die Produktion von Füllpartikeln, Wortwiederholungen, Silbendehnungen oder Korrekturen ursprünglich geplanter Äußerungen [2]. Die Beurteilung des Sprechflusses durch geschulte menschliche Bewerter ist sehr komplex, zeitintensiv und zuweilen inkonsistent. Daher haben de Jong et al. [1] eine Methode vorgestellt, die es mithilfe zweier Praat-Skripte ermöglicht, verschiedene Aspekte des Sprechflusses automatisiert zu messen, z.B. die Artikulationsrate und die Anzahl an Füllpartikeln wie "äh". In [1] wurde die Nutzbarkeit der Skripte für erwachsene Sprecher im Niederländischen und Englischen nachgewiesen.

Bisher wurden Disfluenzen in der Spontansprache von Kindern wenig untersucht, insbesondere nicht zur deutschen Sprache. Vor allem bei Kindern im Vorschulalter lag der Fokus der Forschung eher auf der Redeflussstörung Stottern [3]. Letzteres steht im Gegensatz zu Unterbrechungen des Sprechflusses, der typisch für Spontansprache von Sprechern mit als auch ohne Redeflussstörung ist. Gerade bei Kindern könnte eine automatisierte Messung des Sprechflusses als Indikator für Sprachkompetenz aber ein relevantes Mittel darstellen, da Sprachstandserhebungen vor allem im Vorschulalter durchgeführt werden (in Deutschland bundesweit). Damit

sollen alle Kinder auf ein Sprachniveau zum Schuleintritt gebracht werden, mit dem sie die Schulzeit meistern können [4]. Eine Teilautomatisierung dieser Sprachstandserhebungen würde für eine Erhöhung der Objektivität in der Beurteilung und damit einhergehend auch eine bessere Vergleichbarkeit der Ergebnisse sorgen. Zusätzlich würde dadurch das Personal in den Einrichtungen entlastet, sodass sich mehr auf die direkte pädagogische Arbeit mit dem Kind konzentriert werden kann.

Auch der Aspekt der Zweitsprache wurde in diesem Zusammenhang kaum berücksichtigt, obwohl er in der heutigen Zeit deutlich an Relevanz gewonnen hat: 21% aller Kindergartenkinder in Deutschland wuchsen im Jahr 2022 mit Deutsch als Zweitsprache auf, im Jahr 2008 lag dieser Anteil noch bei 16% [5].

Die vorliegende Studie untersucht, inwiefern sich die in der Phonetik verbreiteten Skripte von de Jong et al. [1] zur Beurteilung des Sprachstands bei Kindern eignen - hier in deutscher Sprache - und in welcher Art und Weise die Methodik angepasst werden könnte. Dabei werden auch die sprechflussbetreffenden Faktoren Artikulationsrate, Pausen und Füllpartikeln bei Vorschulkindern untersucht.

2 Daten

Die in dieser Studie verwendeten Sprachdaten stammen von 20 Kindern im Alter zwischen vierhalb und sechs Jahren, die im Rahmen einer Sprachstandserhebung mithilfe eines Serious Game, der Wuschel-App [6], gewonnen wurden. Das Spiel besteht aus einer zusammenhängenden Geschichte mit 28 verschiedenen Spielsituationen, in denen das Kind jeweils auf eine Frage und eine Nachfrage antwortet, um dem Hauptprotagonisten, dem Hund *Wuschel*, weiterzuhelfen und die Geschichte voranzutreiben. Somit liegen pro Kind 56 Audioaufnahmen (ca. 7 Sekunden pro Aufnahme) vor.

Vor jeder Erhebung werden die Erziehungsberechtigten der Kinder darum gebeten, einen Fragebogen auszufüllen, der für die Studie relevante Informationen zum Umfeld und Sprachhintergrund des Kindes erfasst. Dabei werden unter anderem Angaben zum Alter, der/n Muttersprache/n, der Kontaktdauer mit dem Deutschen (falls relevant) und dem Alter des Kindes zum Zeitpunkt des Eintritts in die Kindertageseinrichtung gemacht.

Das gesamte Projekt zur Sprachstandserhebung, die den Daten dieser Studie zu Grunde liegt, hat zum Ziel, die sprachliche Entwicklung von Kindern zu vergleichen. Daher liegen zu jedem Kind Daten zweier Erhebungszeitpunkte vor (mit einem halben Jahr Abstand).

Hier untersucht wurden Daten von einem der Erhebungszeitpunkte von 20 Kindern aus 10 Kindertageseinrichtungen, wovon 10 Kinder mit Deutsch als Muttersprache (L1) und 10 Kinder mit Deutsch als Zweitsprache (L2) aufwachsen. Jeweils die Hälfte der Kinder beider Gruppen besuchen eine Einrichtung mit Sprachschwerpunkt (mit Unterstützung durch eine geschulte Sprachförderkraft) und die andere Hälfte eine ohne Sprachschwerpunkt.

3 Methode

Als Grundlage der Untersuchung diente die aktuelle Version des Praat-Skripts (v3) zur Erkennung von Silbennuklei, gepaart mit dem dazugehörigen Skript zum Messen von Füllpartikeln von de Jong et al. [1]. Die Skripte erzeugen mithilfe von Praat [7] TextGrid-Dateien mit drei Annotationsebenen: 1) Silbennuklei, 2) Phrasen (inter-pause intervals) und 3) Füllpartikeln.

Für die manuelle Annotation wurden die Ergebnisse der automatischen Detektion auditiv und visuell überprüft und gegebenenfalls korrigiert. Ergänzt wurde die automatische Annotation auf Ebene 2 durch manuell erfasste Pausen, die den Sprechfluss unterbrechen (disfluent) oder nicht unterbrechen (nicht disfluent). Zudem wurde bei der Annotation der Pausen unterschieden,

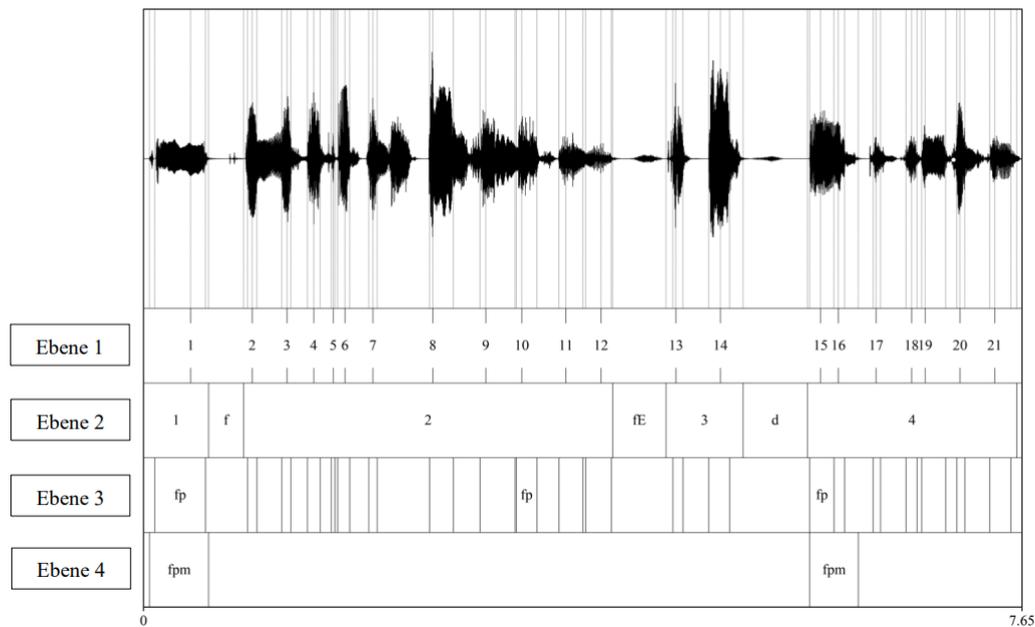


Abbildung 1 – Beispiel einer TextGrid-Datei mit den skriptbasierten Annotationen auf Ebene 1: Silbennuklei, Ebene 2: Pausen und Sprechphasen, Ebene 3: Füllpartikeln. Manuelle Ergänzungen betreffen die Pausentypen auf Ebene 2 und Füllpartikeln auf Ebene 4.

an welcher Position in der Äußerung diese auftreten. So wurden Pausen nach der Frage durch das Spiel und vor der Antwort des Kindes sowie Pausen zwischen zwei Äußerungen des Kindes innerhalb einer Spielsituation gesondert markiert (siehe Tabelle 1). Zusätzlich werden auf einer neuen Ebene (4) *manuell* bestimmte Füllpartikeln annotiert (siehe Abbildung 1).

Tabelle 1 – Annotationsschema von manuell erfassten Pausen bez. des Eindrucks, ob der Sprechfluss als unterbrochen ([d]isfluent) oder als nicht unterbrochen ([f]luent = nicht disfluent) gewertet wurde

Label	Sprechfluss	Beschreibung
f	nicht unterbrochen	Innerhalb einer Äußerung des Kindes
d	unterbrochen	Innerhalb einer Äußerung des Kindes
lf	nicht unterbrochen	Vor Antwort des Kindes
ld	unterbrochen	Vor Antwort des Kindes
fE	nicht unterbrochen	Zwischen zwei Äußerungen des Kindes
dE	unterbrochen	Zwischen zwei Äußerungen des Kindes

Durch die menschlich ergänzten Annotationen lassen sich die automatisch und die manuell bestimmten Pausen, Sprechphasen und Füllpartikeln miteinander vergleichen und die Performance der Skripte als Accuracy- und Precision-Werte errechnen. Zudem wurden die Dauern der disfluenten und der nicht disfluenten Pausen untersucht.

4 Ergebnisse

4.1 Artikulationsrate

Die Übereinstimmung zwischen der automatisch und der manuell bestimmten Artikulationsrate (via Anzahl der Silbennuklei pro Sekunde) ist nicht extrem, aber ausreichend hoch ($\kappa = 0.63$; $p < 0.001$). Das Skript eignet sich somit zu einer ansatzweise genauen Ermittlung der Artikulationsrate, die Aufschluss über den Sprechfluss der Kinder geben könnte. In den Daten wurde auch die Sprechgeschwindigkeit ermittelt, die sich im Gegensatz zur Artikulationsrate auf die

Gesamtsprechzeit bezieht, also *inklusive* Pausen. Die manuelle Messung ergab einen Mittelwert von 1,47 Silben pro Sekunde, mit den automatisierten Analysen durch das Skript wurde eine Sprechgeschwindigkeit von 1,60 Silben pro Sekunde festgestellt.

4.2 Pausen

Die hier angewandte Methode zur Pausenerkennung auf Grundlage der durch das Skript markierten Phrasen scheint auch zur Verwendung bei Daten in deutscher Kindersprache geeignet zu sein. Von den 1.618 skriptbasierten Pausen traten 72 Fälle auf, die nach menschlicher Annotation keine Pausen waren. Damit erreicht die Methode auf den hier verwendeten Daten eine Precision von 0,96.

Nach Annotation der verschiedenen Pausentypen tritt der Pausentyp *If* mit 611 Fällen am häufigsten auf (siehe Tabelle 2). Stellt man die Pausen innerhalb von Äußerungen gegenüber, so tritt in 490 Fällen eine disfluente Pause (*d*), in 191 Fällen eine nicht disfluente Pause (*f*) auf. Bei den Pausen zwischen zwei Äußerungen sind 15 Pausen disfluent (*dE*), gegenüber 141 nicht disfluenten Pausen (*fE*). In einigen Fällen kam es vor, dass das Kind mit seiner Antwort begann, bevor die Frage durch das Spiel fertig gestellt wurde. Dementsprechend existiert nicht in allen 1.120 Aufnahmen eine Pause des Typus *If* oder *Id*. Bei Betrachtung der durchschnittlichen Pausendauern findet sich eine sichtbare Tendenz, dass disfluente Pausen länger sind als nicht disfluente Pausen (im Schnitt 817 ms vs. 383 ms). Diese Tendenz bestätigt sich nach Durchführung eines t-Tests für unabhängige Stichproben: Die Mittelwerte der beiden Kategorien unterscheiden sich signifikant voneinander ($t(531,82) = 16,36$, $p < 0,001$). Ähnlich groß ist die Differenz der durchschnittlichen Pausenlängen bei den Pausen zwischen zwei Äußerungen des Kindes. Hier weisen disfluente Pausen im Schnitt eine Dauer von 2.528 ms und nicht disfluente Pausen eine Dauer von 1.264 ms auf. Auch in diesem Fall kann der t-Test die sichtbare Tendenz als signifikant bestätigen ($t(14,74) = 2,60$, $p = 0,020$).

Tabelle 2 – Vorkommenshäufigkeiten und Mittelwerte (Standardabweichung) der Dauern der verschiedenen Pausentypen

Pausentyp	Häufigkeit	Dauer (SD) in ms
f	191	383 (77)
d	490	817 (574)
If	611	1.073 (864)
Id	98	3.286 (1.656)
fE	141	1.264 (924)
dE	15	2.528 (1.861)

Der Unterschied zwischen den Dauern von disfluenten und nicht disfluenten Pausen wird nochmals größer bei Pausen vor dem Äußerungsbeginn. Eine durchschnittliche Dauer von 3.286 ms bei disfluenten Pausen steht einer Dauer von 1.073 ms bei nicht disfluenten Pausen vor Äußerungsbeginn gegenüber. Der t-Test bestätigt die Signifikanz des Unterschieds ($t(105,61) = 12,95$, $p < 0,001$).

Wie zu erwarten sind Pausen vor oder nach einer Äußerung länger als innerhalb einer Äußerung. Auch die Streuung in der Pausendauer außerhalb einer Äußerung ist größer als innerhalb einer Äußerung (siehe Abbildung 2).

Im Hinblick auf den Sprachhintergrund der Kinder findet sich bei L1-Kindern eine durchschnittliche Dauer von 658 ms bei Pausen innerhalb und 1.284 ms bei Pausen außerhalb einer Äußerung. Bei L2-Kindern sind diese Dauern höher (721 ms und 1.468 ms), der Unterschied zwischen den beiden Spracherwerbstypen erweist sich jedoch ausschließlich für Pausen au-

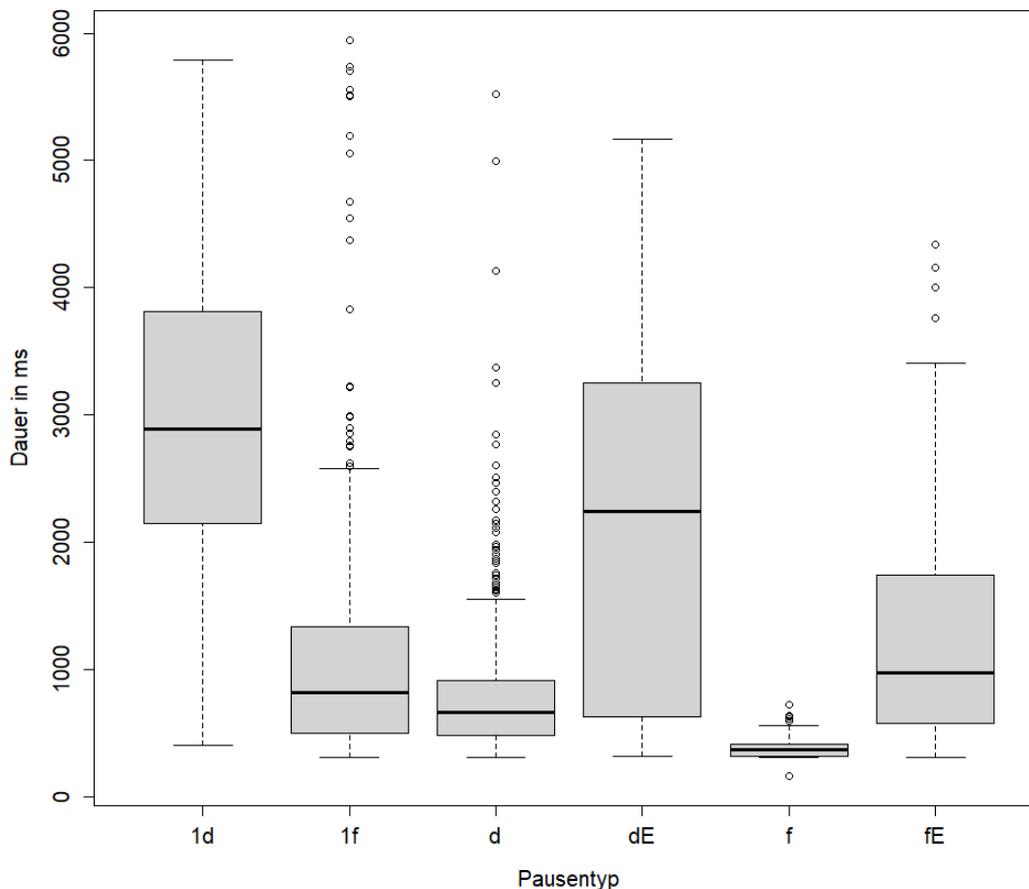


Abbildung 2 – Verteilung der gemessenen Pausenzeiten nach den annotierten Pausentypen.

ßerhalb einer Äußerung als signifikant ($t(678,40) = -1,63$, $p = 0,104$ für Pausen innerhalb und $t(862,89) = -2,19$, $p = 0,029$ für Pausen außerhalb). Kinder, die eine Kita mit Sprachschwerpunkt besuchen, weisen im Schnitt niedrigere Pausendauern auf als Kinder, die eine Einrichtung ohne Sprachschwerpunkt besuchen (633 ms vs. 738 ms bei Pausen innerhalb von Äußerungen, 1.244 ms vs. 1.462 ms außerhalb von Äußerungen). Auch dieser Unterschied ist bei Pausen innerhalb von Äußerungen nicht signifikant, bei Pausen außerhalb von Äußerungen hingegen schon ($t(678,4) = -1,63$, $p = 0,104$ und $t(693,74) = -2,52$, $p = 0,012$). Hinsichtlich Alter und Kontaktdauer der Kinder konnte kein Zusammenhang zu der durchschnittlichen Pausendauer festgestellt werden.

4.3 Füllpartikeln

In den vorliegenden Daten wurden 122 Füllpartikeln manuell annotiert. 47 davon entfallen auf L1-Kinder, 75 auf L2-Kinder. Auffällig ist dabei, dass die Zahl der Füllpartikeln von Kind zu Kind stark unterschiedlich ist (Mittelwert: 6,78; Standardabweichung: 8,25). Ein signifikanter Zusammenhang mit einer der personenbezogenen Variablen (Spracherwerbstyp, Alter, Kontaktdauer, Kita mit Sprachschwerpunkt) konnte allerdings nicht festgestellt werden.

Hinsichtlich der automatisierten Füllpartikelerkennung erreicht das Skript annähernd hohe Werte wie in [1]. Zwar sind die Werte im Allgemeinen etwas niedriger, jedoch bleibt ein Zusammenhang datenübergreifend bestehen: einer vergleichsweise hohen Accuracy ($\sim 0,75$) steht eine geringe Precision ($\sim 0,03$) gegenüber. Dieser liegt zum einen darin begründet, dass beim Sprechen wesentlich weniger Silben als Teil von Füllpartikeln auftreten als Silben lexikalischer

Wörter. Zum anderen tendiert das Skript dazu, viele Silben als Füllpartikeln zu markieren, obwohl es keine sind. Es gibt folglich wesentlich mehr *False Positives* als *False Negatives* (siehe Tabelle 3).

Tabelle 3 – Evaluationsergebnisse des Skripts zur Erkennung von Füllpartikeln.

Precision	Recall	Accuracy	F1
0,03	0,53	0,75	0,05
True Positives	False Positives	True Negatives	False Negatives
65	2.281	7.105	57

Betrachtet man die Umgebung, in dem die automatisch und manuell annotierten Füllpartikeln auftreten, fällt auf, dass die manuell annotierten Füllpartikeln häufig zusammen mit einer Pause auftreten. Von den 122 Vorkommen finden sich 102 unmittelbar vor oder nach einer Pause (± 100 ms Toleranz). Im Gegensatz dazu befinden sich nur 515 der 3.192 (mehrfach durch das Skript als Füllpartikel annotierte Silben wurden in Tabelle 3 auf eine reduziert) durch das Skript als Füllpartikeln annotierten Silben in diesem Pausenkontext.

5 Diskussion

Die Einschätzung des Sprechflusses bei Kindern wird sich stets von der bei Erwachsenen unterscheiden, weil sich Kinder in allen untersuchten Aspekten von erwachsenen Sprechern unterscheiden. Sie reden grundsätzlich langsamer, weshalb die Artikulationsrate nicht in Relation zu der erwachsener Sprecher gestellt werden darf [8]. Deshalb muss eine Beurteilung des Sprechflusses bei Kindern stets unter Berücksichtigung des Alters erfolgen.

Die gemessene Artikulationsrate in der vorliegenden Untersuchung deckt sich weitestgehend mit den Messungen anderer Studien wie die von [8] und [9]. Sie erhöht sich mit zunehmendem Alter der Kinder, liegt aber noch deutlich unter der von Erwachsenen. Dennoch kann auch die Artikulationsrate als Maß für den Sprechfluss bei Kindern dienen, so wie sie bei erwachsenen Sprechern auch einen der Faktoren für das wahrgenommene prosodische Sprachkompetenzlevel darstellt [10].

Die automatisierte Erkennung von Sprechpausen durch das Skript erreicht mit den untersuchten Daten einen hohen Precision-Wert. Somit lassen sich Artikulationsraten und Pausenzeitanteile mit einer soliden Genauigkeit ermitteln, was in eine Sprachstandsmessung ohne großen menschlichen Mehraufwand integriert werden könnte. In den wenigen Fällen, in denen das Skript fälschlicherweise eine Pause erkennt, obwohl tatsächlich keine Pause wahrnehmbar ist, wurde in 74% Fällen ein stark gedehnter Frikativ durch den Sprecher produziert. Lange Frikative scheinen für die Identifizierung von Artikulationsphasen und Sprechpausen ein Problem darzustellen und müssten anders gehandhabt werden, um die Performance weiter zu steigern.

Bei der Untersuchung der Pausendauern fällt auf, dass eine wahrnehmbare Unterbrechung des Sprechflusses von der Dauer der Pausen abhängig sein kann. Dieser Faktor könnte zur automatisierten Bestimmung des Sprechflusses genutzt werden, müsste jedoch stets die Position der Pausen relativ zu den Äußerungen des zu prüfenden Kindes berücksichtigen. Wie oben gezeigt spielt es eine wesentliche Rolle, ob die Pause inner- oder außerhalb der Äußerung platziert ist, um als disfluent wahrgenommen zu werden. Nicht-disfluente Pausen zwischen zwei Äußerungen sind beispielsweise länger als disfluente Pausen innerhalb einer Äußerung. Das Kind hat sozusagen mehr Zeit am Äußerungsbeginn und zwischen zwei Äußerungen, um nicht als disfluent wahrgenommen zu werden.

Bei Pausen vor und nach abgeschlossenen Äußerungen wirken sich etliche Faktoren auf die

Dauer aus, die nicht an den Sprechfluss der Kinder gekoppelt sind. Dazu gehören der Prozess des Verstehens, was von einem verlangt wird, sowie der Mut zur Antwort, aber auch die soziale Verbundenheit, die zwischen den Gesprächspartnern besteht [11]. Somit sollte eine automatisierte Messung der Dauer von Sprechpausen auf solche innerhalb von Äußerungen beschränkt bleiben. Ein solches Verfahren könnte problemlos in eine Sprachstandserhebung integriert werden, um Aussagen über den Sprechfluss und somit die Sprachkompetenz zu treffen.

Im Gegensatz zur Pausenerkennung liefert die automatische Detektion von Füllpartikeln mithilfe der Praat-Skripte keine Ergebnisse, die die menschliche Beurteilung ersetzen können. Zwar wird ein hoher Wert in der Accuracy erreicht, was jedoch damit zu begründen ist, dass wesentlich weniger Silben als Teil von Füllpartikeln auftreten als Silben lexikalischer Wörter, die häufig korrekterweise nicht als Füllpartikeln markiert werden. In dem hier relevanten Anwendungsfall eignet sich daher ein anderes Maß besser, welches die *False Positives* bestraft, wie beispielsweise Precision. Als Auslöser für die Einbußen im Precision-Wert vermuten de Jong et al. [1] längere, ausgedehntere lexikalische Silben, die fälschlicherweise als Füllpartikel klassifiziert werden.

Um weniger *False Positives* zu erhalten, könnte man Anpassungen am Skript vornehmen, wie z.B. die Erhöhung des Schwellenwertes zur Detektion von Füllpartikeln. Jedoch erhöhte sich dadurch die Anzahl an *False Negatives*, da das Skript infolgedessen weniger sensitiv wird und einige tatsächlich auftretende Füllpartikeln nicht mehr als solche erkennt. Ein vielversprechender Ansatz, der sowohl die Anzahl an *True Positives* erhöhen, als auch die Zahl an *False Negatives* niedrig halten würde, könnte zur Erkennung von Füllpartikeln die automatisierte Pausenerkennung miteinbeziehen. Häufig traten in den hier vorliegenden Daten (manuell annotierte) Füllpartikeln in Verbindung mit den mithilfe des Skripts bestimmten Pausen auf. Diese könnten somit als unterstützender Indikator bei der Bestimmung von Füllpartikeln miteinbezogen werden. Möchte das Skript in seiner aktuellen Version eine Silbe als Füllpartikel markieren, könnte man beispielsweise als zusätzliche Zwischenprüfung abfragen, ob sich die Silbe in unmittelbarer Umgebung einer Pause befindet und somit die Anzahl an *False Positives* reduzieren.

Betrachtet man andere Ansätze als denjenigen von den hier benutzten Skripten, z.B. von Reichel et al. [12], so werden Recall- und Precision-Werte bei der Detektion von Füllpartikeln berichtet, die höher liegen als die hier beobachteten. Allerdings bleibt das Grundproblem bestehen, dass ein verlässlich hoher Wert für die Precision noch nicht erreicht wird. Ebenso scheint bei vielen Ansätzen ein hoher Recall-Wert mit einem niedrigen Precision-Wert einherzugehen. Somit besteht ein klar erkennbares Desiderat für zukünftige Forschung.

6 Schlussfolgerungen

Abschließend lässt sich festhalten, dass die Unterstützung bei der Beurteilung des Sprechflusses durch die Praat-Skripte von de Jong et al. [1] auch bei Kindern und in deutscher Sprache eine sinnvolle Option darstellt. Auf eine menschliche Gegenprüfung, gerade bei der Erkennung von Füllpartikeln, sollte jedoch nicht verzichtet werden. Somit könnte eine halbautomatische Bestimmung des Sprechflusses die in Deutschland weit verbreiteten Sprachstandsmessungen bei Vorschulkindern [4] sinnvoll ergänzen.

Danksagung:

Vielen Dank an Dennis Königsmark für seine Hilfe und Unterstützung beim Annotationsprozess und Bernd Möbius für wertvolle Kommentare beim Erstellen des Aufsatzes.

Literatur

- [1] DE JONG, N. H., J. PACILLY, und W. HEEREN: *Praat scripts to measure speed fluency and breakdown fluency in speech automatically*. *Assessment in Education: Principles, Policy & Practice*, 28(4), S. 456–476, 2021. doi:10.1080/0969594X.2021.1951162. URL <https://doi.org/10.1080/0969594X.2021.1951162>. <https://doi.org/10.1080/0969594X.2021.1951162>.
- [2] LICKLEY, R. J.: *Fluency and Disfluency*, Kap. 20, S. 445–474. John Wiley Sons, Ltd, 2015. doi:<https://doi.org/10.1002/9781118584156.ch20>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118584156.ch20>.
- [3] TUMANOVA, V., E. CONTURE, E. LAMBERT, und T. WALDEN: *Speech disfluencies of preschool-age children who do and do not stutter*. *Journal of Communication Disorders*, 49, 2014. doi:10.1016/j.jcomdis.2014.01.003.
- [4] LISKER, A.: *Sprachstandsfeststellung und Sprachförderung im Kindergarten sowie beim Übergang in die Schule*. *Expertise im Auftrag des Deutschen Jugendinstituts.*, 2010. URL http://www.dji.de/bibs/Expertise_Sprachstandserhebung_Lisker_2010.pdf.
- [5] MAAZ, K., C. ARTELT, P. BRUGGER, S. BUCHHOLZ, S. KÜHNE, H. LEERHOFF, T. RAUSCHENBACH, J. SCHRADER, und S. SEEBER: *Bildung in Deutschland 2022: Ein indikatorengestützter Bericht mit einer Analyse zum Bildungspersonal*. wbv Publikation, 2022.
- [6] ROCHE, J., S. HABERZETTL, G. PAGONIS, M. JESSEN, und N. WEIDINGER: *Serious Games in der Sprachstandsermittlung*, S. 340–358. Narr Francke Attempto Verlag, 2019. doi:<http://dx.doi.org/10.22028/D291-35846>.
- [7] BOERSMA, P. und D. WEENINK: *Praat: doing phonetics by computer (version 6.4.04)*. 2024. URL <http://www.praat.org>.
- [8] MAHR, T. J., J. U. SORIANO, P. J. RATHOUZ, und K. C. HUSTAD: *Speech development between 30 and 119 months in typical children ii: Articulation rate growth curves*. *Journal of Speech, Language, and Hearing Research*, 64(11), S. 4057–4070, 2021. doi:10.1044/2021_JSLHR-21-00206.
- [9] WALKER, J. F., L. M. D. ARCHIBALD, S. R. CHERNIAK, und V. G. FISH: *Articulation rate in 3- and 5-year-old children*. *Journal of Speech, Language, and Hearing Research*, 35(1), S. 4–13, 1992. doi:10.1044/jshr.3501.04.
- [10] KALLIO, H.: *The prosody underlying spoken language proficiency. Cross-lingual investigation of non-native fluency and syllable prominence*. Ph.D. thesis, 2022. doi:10.13140/RG.2.2.29682.99524.
- [11] TEMPLETON, E., L. CHANG, E. REYNOLDS, M. LEBEAUMONT, und T. WHEATLEY: *Fast response times signal social connection in conversation*. *Proceedings of the National Academy of Sciences*, 119, S. e2116915119, 2022. doi:10.1073/pnas.2116915119.
- [12] REICHEL, U. D., B. WEISS, und T. MICHAEL: *Filled pause detection by prosodic discontinuity features*. In P. BIRKHOLZ und S. STONE (Hrsg.), *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2019*, S. 272–279. TUDpress, Dresden, 2019. URL https://www.essv.de/pdf/pdf/2019_272_279.pdf.