

INDIVIDUELLE AUSPRÄGUNG VON ATMUNGSPAUSEN IN DER MUTTER- UND IN DER FREMDSPRACHE ALS ANZEICHEN KOGNITIVER BELASTUNG

Jürgen Trouvain und Bernd Möbius

Universität des Saarlandes, Computerlinguistik und Phonetik

Email: trouvain [at] coli.uni-saarland.de

Abstract: Die vorliegende Studie analysiert die Häufigkeit und Dauer von Atmungsgeräuschen sowie ihren proportionalen Anteil an Sprechpausen in mutter- und fremdsprachlichen Produktionen deutscher und französischer Sprecher. Getestet wurde die Hypothese, dass die durch größeren Planungsaufwand erhöhte kognitiv-linguistische Belastung in der L2 zu einer Steigerung der Atmungsaktivität führt. Die Ergebnisse bestätigen die Erwartung tendenziell, weisen aber auch auf eine ausgeprägte individuelle Verwendung des Pausenverhaltens hin. Außerdem unterscheiden sich die Sprecher der beiden Sprachen als Gruppe in ihrem Pausen-Timing, unabhängig davon, ob L1- oder L2-Äußerungen produziert wurden.

1 Einführung

Einatmungsgeräusche in Sprechpausen sind phonetische Erscheinungen, die bekanntermaßen in natürlicher Sprache vorkommen, aber bislang kaum in der Sprachsynthese, der automatischen Spracherkennung und der automatischen Sprecheridentifizierung genutzt werden.

Ziel der vorliegenden Studie ist, zu überprüfen, ob und inwieweit sich Sprecher bei muttersprachlichen (L1) Texten bezüglich Sprechpausen mit Einatmungsgeräuschen anders verhalten als in entsprechenden fremdsprachlichen (L2) Texten. Der erhöhte Planungsaufwand in der L2 führt zu erhöhter kognitiver Belastung [5]. Zusätzlich kann ein erhöhter Grad an Stress eine verstärkte Einatmung bewirken. Daher ist davon auszugehen, dass in L2-Sprache die *absoluten* Werte für Anzahl und Dauer von Pausen deutlich höher sind als in L1-Sprache. Dies gilt ebenso für die Anzahl von Unflüssigkeiten wie Satzabbrüchen, gefüllten Pausen und Reparaturen ("repair fluency") und die Artikulationsrate.

Ungeklärt hingegen ist das Atmungsverhalten in der L2 *relativ* zur L1. Eine Erhöhung an Pausierungsaktivität muss nicht unbedingt zu einer Erhöhung der Atmungsaktivität im selben Zeitrahmen führen. Jedoch wäre zu erwarten, dass eine stärkere kognitive Belastung bei der L2-Produktion zu mehr Einatmung oder zu längerer Einatmung führt, da diese Anzeichen verstärkter Planung bevorstehender Sprechabschnitte darstellen [1], wobei diese durch Dauer, aber nicht durch Intensität reflektiert wird [6].

Gleichzeitig gibt es auch Befunde, die eine starke individuelle Ausprägung des Atmungsverhaltens (innerhalb der Muttersprache) belegen [4]. Unklar bleibt aber, inwieweit sich individuelle Muster in der gewohnten Sprache (meist L1) über Kommunikationssituationen und Textsorten hinweg als konstant erweisen und inwieweit dies auch in der ungewohnten Sprache (meist L2) der Fall ist.

1.1 Hypothesen

Aus den vorangehenden Ausführungen lassen sich folgende Hypothesen ableiten:

- In der L2 werden generell mehr Pausen sowie längere Pausen produziert als in der L1. Der Gesamtanteil der Pausendauern müsste daher in der L2 auch größer sein als in der entsprechenden L1.

- Die mittlere Dauer der Pausen – sowohl der Pausen mit als auch solcher ohne hörbare Einatmung – sollte dementsprechend in der L2 auch länger sein.
- Da Sprechunflüssigkeiten nicht geplant werden, sollten sie auch nicht an einer Pause mit Einatmung stattfinden.

Über bestimmte individuelle Ausprägungen kann keine Hypothese aufgestellt werden. Daher ist es in der vorliegenden Studie das Ziel, das Ausmaß solcher Ausprägungen festzustellen und gegebenenfalls bestimmte Muster zu erkennen und zu erklären.

2 Experiment

2.1 Methode

Aus einem im Entstehen begriffenen zweisprachigen Parallelkorpus [8] wurden jeweils fünf Anfänger (Niveau A1/A2 im Europäischen Referenzrahmen CEFR) mit Französisch bzw. mit Deutsch als L1 in der Teilaufgabe "Vorlesen von Texten" untersucht. Beim hier untersuchten Text handelt es sich um die bekannte Geschichte "Die drei kleinen Schweinchen"/"Les trois petits cochons", die in beiden Sprachen vorliegt und von allen zehn Versuchspersonen (Vpn) in ihrer L1 und in ihrer L2 in einer ruhigen Büroumgebung vorgelesen und aufgenommen wurde.

Aus den ersten 30 Sekunden der insgesamt 20 Aufnahmen wurden die Pausen (eingeteilt in Atempausen und Nicht-Atempausen) bezüglich ihrer Anzahl, ihrer Dauer und der Dauer der jeweiligen Einatmungsgeräusche analysiert. Ein Einatmungsgeräusch lag dann vor, wenn es als hörbar eingestuft wurde. Die entsprechende Dauer wurde durch sichtbare Intensitätsunterschiede im Oszillogramm und vor allem im Spektrogramm bestimmt (vgl. Abbildung 1). Wenn bei Pausen auch Merkmale von Sprechunflüssigkeiten aufgetreten sind wie Silbendehnungen (aber nicht Dehnungen von Stille), so wurden diese eigens vermerkt.

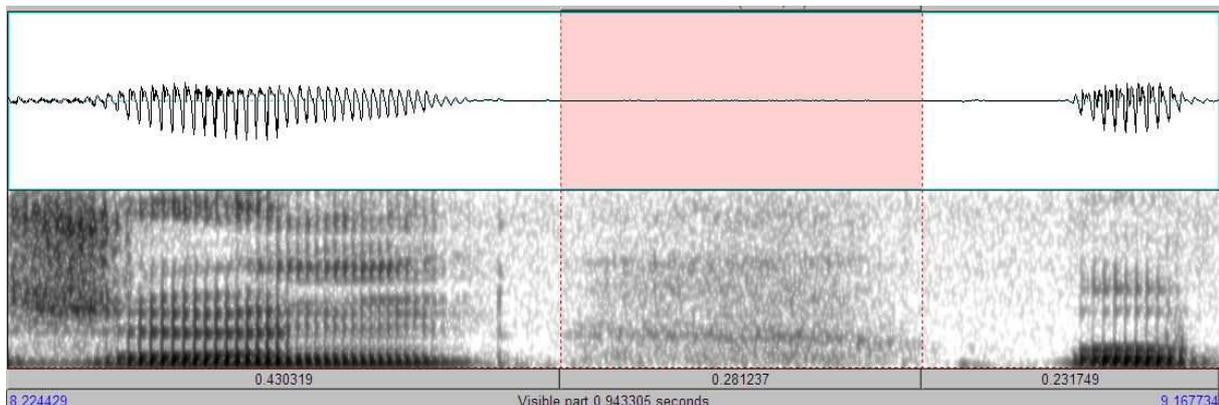


Abbildung 1 – Oszillogramm und Spektrogramm (0-8 kHz) einer Atempause mit linkem und rechten Kontext ("... jeunes – Atempause – où ...") eines deutschen Sprechers im L2-Text. Eingefärbt ist das hörbare und im Spektrogramm sichtbare Einatmungsgeräusch. Die stillen Anteile der Atempause vor und nach dem Geräusch sind hier nicht gesondert markiert.

2.2 Ergebnisse

2.2.1 Anzahl der Pausen

Wie in Abbildung 2 ersichtlich zeigt eine leichte Mehrheit der Sprecher in ihrer L2 mehr Pausen (n=6), wobei Atempausen und Nicht-Atempausen zusammen zählen. Zwei Vpn produzieren in L1 und L2 genau so viele Pausen und bei zwei Vpn gibt es in der L1 mehr Pausen.

Ähnlich verhält es sich mit der Anzahl an Atempausen. Die Hälfte der Vpn produziert in der L2 mehr Atempausen als in der L1, zwei Vpn gleich viel Atempausen in L1 und L2 und zwei Vpn haben weniger Atempausen in der L2 als in der L1.

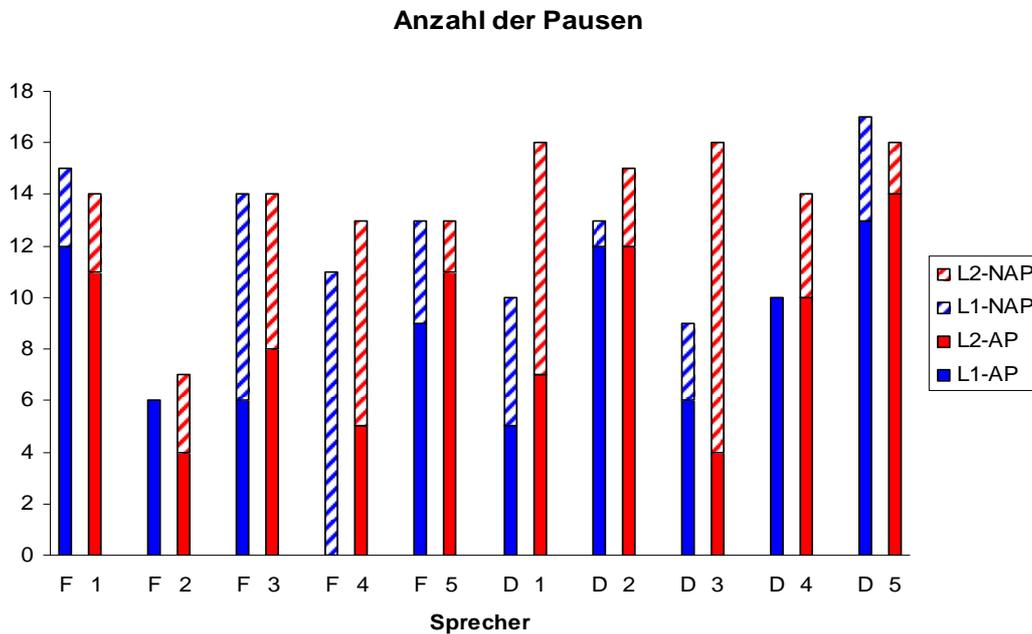


Abbildung 2 – Anzahl der Pausen innerhalb der ersten 30 Sekunden pro Sprecher. Der jeweils linke Balken bezieht sich auf das Vorlesen in der Muttersprache (L1), der jeweils rechte Balken auf die Fremdsprache (L2). Innerhalb der Balken sind die Pausen in Atemspausen (AP) und Nicht-Atemspausen (NAP) unterteilt. "F" und "D" stehen für die jeweilige Muttersprache.

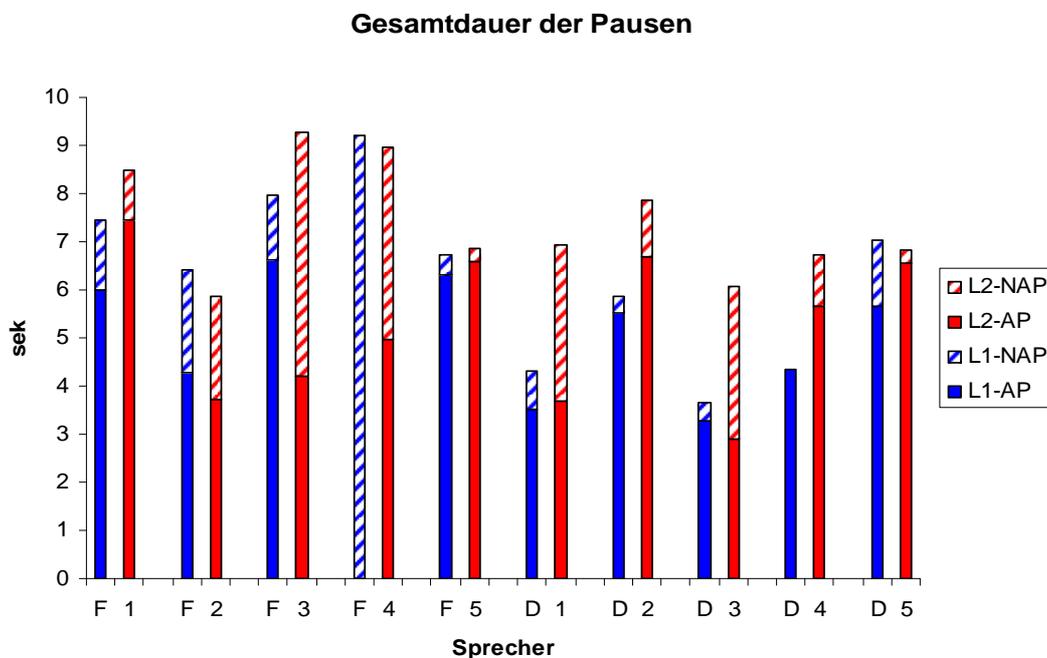


Abbildung 2 – Gesamtdauer der Pausen und der hörbaren Einatmungen innerhalb der ersten 30 Sekunden pro Sprecher. Der jeweils linke Balken bezieht sich auf das Vorlesen in der Muttersprache (L1), der jeweils rechte Balken auf die Fremdsprache (L2). Innerhalb der Balken sind die aufsummierten Pausendauern in hörbare Einatmung (bei Atemspausen), Stille (bei Atemspausen) und Stille bei Nicht-Atemspausen unterteilt. "F" und "D" stehen für die jeweilige Muttersprache.

2.2.2 Gesamtpausenzeit

Bezüglich der Gesamtpausenzeit zeigt eine Mehrheit (n=7) einen größeren Wert in der L2, bei drei Vpn hingegen ist der Wert in der L1 größer (Abbildung 3).

Betrachtet man nun lediglich die aufsummierte Gesamtzeit der Atempausen so ist das Verhältnis gleich geblieben: 7 Vpn haben eine längere Atempausenzeit in der L2, drei Vpn in ihrer L1. Allerdings handelt es sich nicht immer um die gleichen Individuen, die diesen Mustern entsprechen, z.B. Sprecher D5.

2.2.3 Mittlere Dauer der Pausen und Einatmungsgeräusche

Bei der mittleren Dauer von Atempausen und Einatmungsgeräuschen zeigen sich Unterschiede zwischen den Individuen als auch zwischen den beiden L1-Gruppen. In Tabelle 1 kann man an den Mittelwerten für Atem- und Nichtatempausen erkennen, dass die D-Gruppe (mit Deutsch als Muttersprache) sowohl in ihrer L1 als auch in ihrer L2 kürzere Pausen als die F-Gruppe (mit Französisch als L1) produziert. Das heißt also auch, dass die Deutschsprachigen in ihren französischen Versionen im Schnitt kürzere Pausen aufzeigen als die französischen Muttersprachler.

Eine Konstanz über L1 und L2 hinweg ist bei der durchschnittlichen Dauer der Einatmungsgeräusche zu erkennen. Die F-Gruppe zeigt einen Wert von 332 ms gegenüber 300 ms bei der D-Gruppe.

	F1-F5		D1-D5	
	L1	L2	L1	L2
Nicht-Atempausen	523 (521)	432 (322)	246 (173)	297 (252)
Atempausen	706 (188)	747 (393)	485 (204)	546 (213)
Stille	374 (174)	414 (305)	180 (107)	248 (160)
Atmungsgeräusch	331 (90)	333 (151)	301 (144)	298 (107)

Tabelle 1 – Mittelwerte (Standardabweichung in Klammern) der Dauern der Nicht-Atempausen und der Atempausen, letztere nochmals unterteilt in Stille und Atmungsgeräusch

Die individuelle Verteilung der Dauern der Atempausen sowie ihrer Stille- und Geräusch-Anteile ist in Abbildung 4 dargestellt. Im Vergleich zu den Stille-Anteilen sind die Werte für die Einatmungsgeräusche bei fast allen Sprechern relativ konstant. Ausnahmen dazu bilden die Sprecher F2 (in seiner L2) und D1 (in seiner L1), F4 fällt wegen seiner fehlenden hörbaren Atmung bei L1 hier weg.

Generell gilt, dass der Stille-Anteil in der L2 länger ist als in der L1. Nur bei zwei Vpn ist der Stille-Anteil in der L2 geringer als in der L1 (F3, F5).

Etwas größere individuelle Unterschiede lassen sich bei den Dauern der Nicht-Atempausen feststellen (Abbildung 5). Die Tendenz hierbei ist, dass in der L2 diese Pausenart gedehnt wird. Für drei Vpn (F4, F5, D5) trifft dieser Befund nicht zu.

Mittelwerte Stille, hörbare Einatmung bei Atempausen

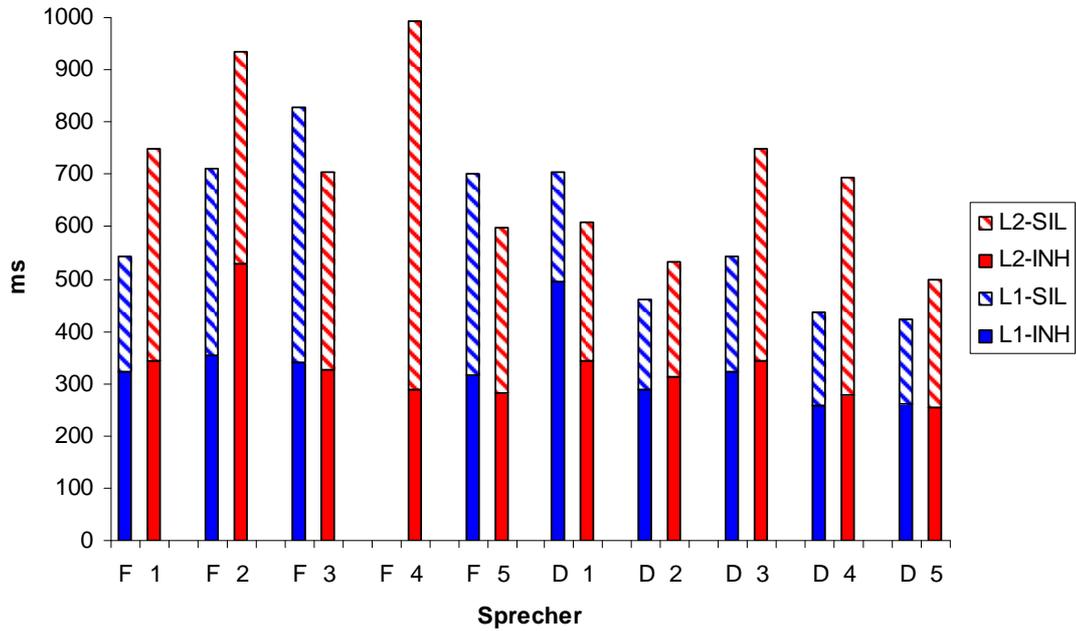


Abbildung 4 – Mittlere Dauer der Stille-Anteile (SIL) und der hörbaren Einatmungen (INH) bei Atempausen innerhalb der ersten 30 Sekunden pro Sprecher. Alle anderen Angaben sind wie bei den vorhergehenden Abbildungen.

Mittelwerte der Nicht-Atempausen

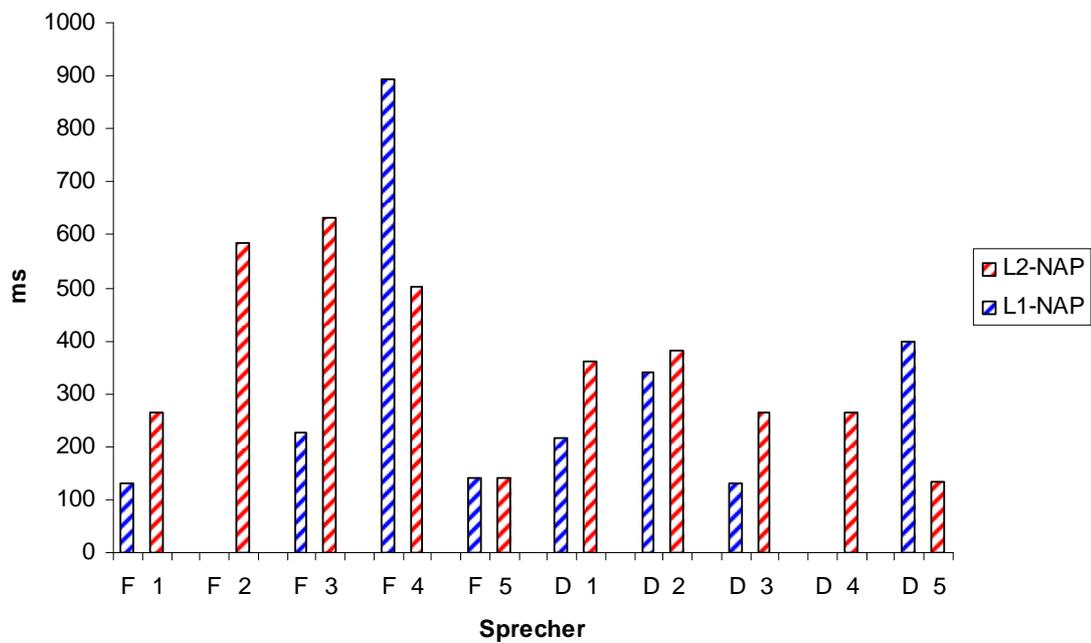


Abbildung 5 – Mittlere Dauer der Nicht-Atempausen innerhalb der ersten 30 Sekunden pro Sprecher. Alle anderen Angaben sind wie bei den vorhergehenden Abbildungen.

2.2.4 Pausen und Sprechunflüssigkeiten

Darüber hinaus gibt es einzelne Pausen, die zum Eindruck des unflüssigen Sprechens führen, meist verbunden mit einer Lautdehnung oder der Wiederholung eines Wortes bzw. einer Wortverbindung. Diese Sprechflusspausen sind bei acht der zehn Sprecher zu beobachten. 15 der insgesamt 19 Sprechflusspausen traten in der L2 auf, vier in der L1. Zumeist betrifft es Nicht-Atempausen (n=16), Atempausen hingegen lediglich drei Mal, wobei es sich um extrem lange Atempausen (minimal 970 ms) handelt.

2.3 Diskussion

Die einzelnen Faktoren der erwarteten Tendenzen "mehr Pausen, längere Pausen, mehr Atempausen, daher auch mehr Atmungszeit" in der L2 gegenüber der L1 sind zwar mehrheitlich festzustellen, dafür aber nur schwach ausgeprägt. Die Annahme, dass die Mehrheit der Sprecher all diesen Faktoren in der erwarteten Tendenz folgen, ist nicht bestätigt worden. Genau betrachtet handelt es sich um nur einen (von zehn) Sprechern, D1, der den Erwartungen bezüglich Anzahl und Gesamtdauer der Pausen und Atempausen in L1 gegenüber L2 vollständig entspricht.

In Einklang mit älterer Literatur wie z.B. Goldman Eisler [2] kann man feststellen, dass die individuelle Variation sehr stark ausgeprägt ist. Beispielsweise zeigt Vp F4 ein spezielles Verhalten, indem sie ausschließlich Nicht-Atempausen (d.h. Pausen ohne *hörbare* Einatmung) in der L1 benutzt. Im völligen Gegensatz dazu ist bei den Sprechern F2 und D4 jede Pause mit einem Atemgeräusch versehen. Manche Vpn verhalten sich wie vorhergesagt mit längerer Einatmung und längeren Atmungspausen in der kognitiv schwierigeren L2, andere hingegen verkürzen die Atmungspausen und teilweise auch ihr Einatmungsgeräusch bei gleichzeitiger Erhöhung der Anzahl der Pausen.

Die interessante Beobachtung gänzlich fehlender *hörbarer* Einatmung bei Sprecher F4 wirft die Frage auf, wie häufig Atemtätigkeiten ohne *visuell* erkennbares akustisches Korrelat bzw. ohne auditive Wahrnehmung bleiben. Es ist davon auszugehen, dass zur Einatmung kinematische Bewegung erforderlich ist und daher auch bei Sprecher F4 vorhanden gewesen sein muss. Hier könnten atmungskinematische Aufzeichnungen bei gleichzeitigen akustischen Aufnahmen über diesen Punkt mehr Aufschluss geben. Dies könnte auch zur Klärung der Frage beitragen, ob und wie sich Ruheatmung von Sprechatmung akustisch unterscheidet. Auf jeden Fall darf man davon ausgehen, dass bei den restlichen neun Vpn eine unbekannte Anzahl von stattgefundenen Einatmungszyklen akustisch keinen Niederschlag gefunden hat und diese daher zur Gruppe der "Nicht-Atmungspausen" gezählt wurden.

Die Dauer des Einatmungsgeräusches (wenn es denn festgestellt wurde) ist erstaunlich konstant über Sprecher und Sprachen hinweg. In Atempausen sind es vor allem die Stillephasen, die flexibel zur Dehnung der Pausen eingesetzt werden und als Korrelate der erhöhten kognitiven Belastung in Frage kommen.

3 Schlussfolgerungen

Globale Parameter des Pausenverhaltens (Gesamtanzahl der Pausen, Gesamtpausenzeit, mittlere Dauer von Atem- und Nichtatempausen) können als Anzeichen erhöhter kognitiver Belastung beim Vorlesen von L2-Texten betrachtet werden. Jedoch scheint die individuelle Verwendung der Atmung in dieser Bedingung stärker als vermutet ausgeprägt zu sein. Es bleibt zukünftigen Studien überlassen zu zeigen, ob und inwieweit sich diese Individualität auch bei anderen Textsorten beobachten lässt bzw. als wie stabil sich individuelle Sprechpausenmuster über Textsorten hinweg erweisen.

Neben der stark individuellen Ausprägung überrascht der hier dargelegte Befund, dass sich Sprecher der beiden untersuchten Sprachen als Gruppe in ihrem Pausen-Timing in L1 und L2 unterscheiden, sowohl bei den Nicht-Atempausen als auch bei den stillen Anteilen von Atempausen und den feststellbaren Geräuschanteilen darin. Die Fortführung der Aufnahmen und ihrer Annotationen des hier verwendeten zweisprachigen Korpus [8] auf 100 statt 10 Sprecher wird eine zuverlässigere Aussage diesbezüglich erlauben.

Die große Variabilität in der zeitlichen Strukturierung der Pausen könnte einen weiteren Ansatzpunkt bieten, synthetisch erzeugte Sprache mehr der menschlichen Sprachproduktion anzugleichen. Bislang werden in der Sprachsynthese vorgelesene Texte im Allgemeinen mit wenigen statisch zugewiesenen Pausendauern modelliert. Ein weiterer Punkt bei der Modellierung des menschlichen Sprechvorbildes sollte die Verwendung verschiedener Arten von Einatmungsgeräuschen in längeren Texten sein. Eine positive Auswirkung auf das (unbewusste) Empfinden von Natürlichkeit ist bereits bei isolierten Sätzen nachgewiesen worden [10], auch wenn hierbei ein bloßes Hinzufügen eines festen Geräusches nicht unbedingt Erfolg verspricht [9].

Viele Applikationen der Sprachsynthese generieren eher einzelne Sätze als längere Texte. Pausen dürften daher, wenn überhaupt, eine untergeordnete Rolle spielen. Eine differenziertere Pausenmodellierung könnte aber dann zum Tragen kommen, wenn Textlänge, Textsorte und möglicherweise die Individualität des synthetischen Sprechers von dem genannten Standardfall abweichen, z.B. bei individualisierten Sprechprothesen nach Stimmverlust, Erzeugung dramatischer Texte für Blinde [7] oder anderen Sprechstilen [3].

Im Bereich des computergestütztes Fremdsprachenlernens spielt die Erkennung von Pausen, und demzufolge die Zuordnung des Einatmungsgeräusches, eine wesentliche Rolle. Häufig werden tatsächliche Sprechpausen nicht richtig mit den vorhergesagten Pausen aligniert, was unter Umständen auf die durch Einatmung verursachte Friktion zurückzuführen ist. Gesonderte Studien müssen in der Zukunft zeigen, wie stabil die Pausenerkennung auf Grund von eventuell störenden Einatmungsgeräuschen funktioniert und inwieweit sich diese frikativähnlichen Geräusche von akustisch ähnlichen Realisierungen von /h/ und Aspirationsphasen stimmloser Plosive unterscheiden.

Ein letzter Punkt betrifft die Erkennung eines individuellen Sprechers durch Menschen. Können menschliche Hörer Einatmungsgeräusche in Atempausen von verschiedenen Sprechern den zugehörigen sprachlichen Äußerungen verlässlich zuordnen? Diese Fragestellung wäre insbesondere für die forensische Phonetik von Interesse. Sie würde idealerweise ergänzt werden durch eine akustische Untersuchung der Intensität und weiterer akustischer Korrelate der kinematischen Bewegungen der Sprechatmung.

Literatur

- [1] Fuchs, S., Petrone, C., Krivokapić, J. & Hoole, Ph. 2013. Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics* 41, pp. 29-47.
- [2] Goldman Eisler, F. 1968. *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.
- [3] Kruschke, H. 2007. *Prosodische Analyse von Sprechstilen: Ein Beitrag zur Simulation von Sprechstilen in synthetischer Sprache*. (Studententexte zur Sprachkommunikation 45). Dresden: TUDpress.
- [4] Lauf, R. 2001. Aspekte der Sprechatmung: Zur Verteilung, Dauer und Struktur von Atemgeräuschen in abgelesenen Texten. In: Braun, A. (Hg.) *Beiträge zu Linguistik und Phonetik. Festschrift für Joachim Göschel zum 70. Geburtstag (ZDL Beihefte 118)*. Stuttgart: Franz Steiner Verlag, pp. 406-420.
- [5] Mitchell, H.L. Hoit, J.D. & Watson, P.J. 1996. *Cognitive-linguistic demands and speech*

- breathing. *Journal of Speech and Hearing Research* 39, pp. 93-104.
- [6] Rochet-Capellan, A. & Fuchs, S. 2013. The interplay of linguistic structure and breathing in German spontaneous speech. *Proc. Interspeech*, Lyon, pp. 2014-1018
- [7] Rusko, M., Trnka, M., Darjaa, S. & Hamar, J. 2013. The dramatic piece reader for the blind and visually impaired. *Proc. 4th Workshop on Speech and Language Processing for Assistive Technologies*, Grenoble, pp. 83-91.
- [8] Trouvain, J., Laprie, Y., Möbius, B., Andreeva, B., Bonneau, A., Colotte, V., Fauth, C., Fohr, D., Jouviet, D., Mella, O., Jügler, J. & Zimmerer, F. 2013. Designing a bilingual speech corpus for French and German language learners. *Proc. Corpus et Outils en Linguistique, Langues et Parole: Statuts, Usages et Ménuages*, Strasbourg, pp. 32-34.
- [9] Trouvain, J. & Möbius, B. 2013. Einatmungsgeräusche vor synthetisch erzeugten Sätzen – Eine Pilotstudie. *Proc. 24. Konferenz Elektronische Sprachsignalverarbeitung (ESSV '13)*, Bielefeld, pp. 50-55.
- [10] Whalen, D.H., Hoequist, Ch.E. & Sheffert, S. 1995. The effects of breath sounds on the perception of synthetic speech. *Journal of the Acoustical Society of America* 97, pp. 3147-3153.