

# ZUR ANNOTATION NICHT-VERBALER VOKALISIERUNGEN IN KORPORA GESPROCHENER SPRACHE

Jürgen Trouvain<sup>1</sup>, Malte Belz<sup>2</sup>

<sup>1</sup>Universität des Saarlandes, <sup>2</sup>Humboldt-Universität zu Berlin  
trouvain@coli.uni-saarland.de

**Kurzfassung:** Nicht-verbale Vokalisierungen (NVV) wie z. B. Atmungsgeräusche, affektive Äußerungen, Füller oder Clicks sind ein Phänomen, das häufig in gesprochener Sprache zu beobachten ist. Dabei kommen NVV häufiger in dialogischer spontaner Sprache als in monologischer geskripteter Sprache vor. Die vorliegende Studie schlägt eine Mehr-Ebenen-Struktur zur Annotation von NVV vor und diskutiert die Herausforderungen anhand zweier Korpusstichproben und einem (nicht-exhaustiven) Vergleich zur bisherigen Praxis der NVV-Annotation.

## 1 Einführung

Gesprochene Korpora enthalten meistens eine Transliteration, die sich nach einer bestimmten (z. B. orthographischen) Normalisierung richtet. Neben diesen verbalen Vokalisierungen existieren auch eine Reihe nicht-verbaler Vokalisierungen (NVV), die in gesprochener Sprache vorkommen, wie bspw. Lachen und Räuspern, und deren orthographische Wiedergabe problematisch oder gar unmöglich ist [1]. Trouvain und Truong [2] unterscheiden sechs Klassen von NVV, wobei eine klare Trennung zwischen verbalem und non-verbalem Status der jeweiligen Vokalisierung oftmals nicht gegeben ist.

**Vegetative Äußerungen** dienen nicht primär der Kommunikation und nicht alle können bewusst kontrolliert werden. Beispiele sind Geräusche der Ein- und Ausatmung, Schnarchen, Stöhnen, Schluck- und Kaugeräusche, Husten oder Räuspern.

Zu den **affektiven Äußerungen** gehören Lachen, Weinen, Jubeln oder Aufschreien und in einem weiteren Sinne Interjektionswörter wie *jippie* oder *igitt*.

**Interjektionen** werden oftmals als eigene Wortklasse angesehen, obwohl es eine allgemein akzeptierte Definition von Interjektionen nicht gibt. Ausprägungen sind solche mit affektivem Gehalt (z. B. *Mist!*, *Autsch!*), lautmalerischer Natur (z. B. *miau*) oder mit appellativem Charakter (z. B. *pst!*, *sch*), aber auch mit anderen ungewöhnlichen Lauten wie *ts-ts-ts* (drei Zungenschmalzlaute) zum Ausdruck von Missbilligung. Von “gezähmten” Interjektionen mit einem hohen Grad an Konventionalisierung unterscheiden sich “wilde” Interjektionen [3], die nicht oder kaum konventionalisiert sind und sich auch schlecht verschriften lassen, z. B. das Imitieren von Tier- oder Maschinengeräuschen.

**Füller** (oder “gefüllte Pausen”) werden oft mit *äh* oder *ähm* wiedergegeben. In sprachlichen Dialogen beträgt ihre Häufigkeit etwa 1–5 % aller Wörter [4, 5], in aufgabenbasierten Dialogen tendenziell mehr als in aufgabenfreien.

Eine weitere Klasse mit unklarem Wortstatus sind **Feedback-Äußerungen**, häufig auch als “backchannels” oder “Hörersignale” bezeichnet. Sie dienen in Gesprächen zum einen dazu, die Verständlichkeit über den auditiv-akustischen Kanal (“backchannel”) zu signalisieren, zum anderen aber auch dazu, Zustimmung und weitere Kommentare zu äußern, ohne das Rederecht des Gesprächspartners zu beanspruchen. Häufige Formen sind *hm*, *ja* oder *aha*.

**Melodische Äußerungen** gehören zum universellen phonetischen Repertoire. Melodien ohne Text können gesungen, gesummt oder gepfiffen werden.

In Korpora mit gesprochener Sprache erwarten wir melodische Äußerungen sehr selten, Feedback-Äußerungen nur in Dialogen, Füller lediglich in Spontansprache, affektive Äußerungen bis auf Lachen eher selten und vegetative Äußerungen bis auf Atmungsgeräusche auch eher selten. “Gezähmte” Interjektionen betrachten wir hier als verbales Material und daher nicht als NVV. “Wilde” Interjektionen werden in Korpora gesprochener Sprache wohl nicht häufig vorkommen.

NVV können eine Vielzahl von Funktionen erfüllen, von denen hier einige genannt werden. So sind Feedback-Äußerungen in Dialogen unerlässlich. Ihre Kürze und ihre akustische “Zurückgenommenheit” lassen zwei Gesprächspartner zur gleichen Zeit wichtige Information auf verschiedenen Ebenen austauschen: von der Turn-Organisation bis zu Zustimmung und anderen Kommentaren. Füller können in Dialogen eine Rolle bei der Turn-Organisation spielen, aber auch beim Lenken der Aufmerksamkeit des Partners auf wichtige und neue Information. Lachen und andere affektive Äußerungen stellen oftmals wichtige Kommentare in Kurzform dar, die sowohl den inneren Zustand eines Sprechers als auch dessen Einstellungen widerspiegeln. Ebenso erregen auch “wilde” Interjektionen durch ihre ungewöhnliche lautliche Ausprägung die Aufmerksamkeit des Zuhörers. Auch wenn Einatmung und die damit verbundenen Geräusche ursprünglich einer vegetativen Funktion dienen, so können sie in gesprochener Sprache Funktionen in der sprachlichen Kommunikation übernehmen, wie z. B. das Andeuten der Länge der folgenden prosodischen Phrase, die Beanspruchung des Rederechts oder als Anzeichen von Aufgeregtheit.

NVV werden in gesprochenen Korpora uneinheitlich erfasst [6], und der Detailgrad ihrer Annotation hängt naturgemäß stark von der spezifischen Forschungsfrage ab. Daher existiert bislang kein standardisiertes Annotationsschema. Für Fragestellungen, die NVV in ihre Analysen einbeziehen möchten, kann ein Schema zur Annotation der häufigsten Ausprägungen von NVV eine wichtige Orientierungshilfe darstellen. Unser Vorschlag beruht auf einer Analyse von Sprechpausen mit NVV in einem spontanen dialogischen Sprechstil und in laut gelesener Sprache.

## 2 Alternative Annotationsschemata

In Annotationsschemata gesprochener Korpora findet man etliche Parallelen, aber auch Unterschiede. Im Folgenden stellen wir Auszüge aus drei verschiedenen Annotationsschemata vor, soweit sie NVV betreffen. Die Auswahl wurde willkürlich getroffen.

Beim Diapix-Wildcat-Korpus [7] handelt es sich um eine Sammlung annotierter Gesprächsdaten, die aufgabenbasiert erhoben wurden: zwei Versuchspersonen sollen die Unterschiede in ihren jeweils zur Verfügung stehenden Bildern herausfinden. Für NVV kommen zwei Gruppen von Annotationsklassen in Frage: zum einen “Hesitation sounds/filled pauses & yes/no sounds”, bei denen *uh, um, er, ah, eh, oh, mm, hm, yuh, huh* oder Kombinationen wie *uhhm, uhuh, mmhm, mmmm, uhuh, yuhuh*, etc. benutzt werden können. Zum anderen gibt es “non-speech sounds” mit *laughter* (entweder frei stehend oder Teil eines Wortes bzw. über mehrere Wörter hinweg), *breaths & sighs, lip smack, garbage*. Zu der Residualklasse gehören zum einen Geräusche, die nicht vom Sprecher stammen (z. B. Mikrofonploppen), zum anderen auch alle Vokalisierungen des Sprechers, die nicht in eine der vorhergehenden Klassen passt (z. B. Räusperrn).

Bei GAT II [8] handelt es sich um eine Konvention zur Transkription von Gesprächen, die vor allem in der deutschsprachigen konversationsanalytischen Forschung Verwendung findet. Diese Richtlinien sind daher nicht für die automatische Analyse gedacht, sondern stellen

eher eine individuelle Transkription mit einem hohen Grad an Interpretation dar. Nonverbale Handlungen und Ereignisse sind Charakterisierungen parasprachlicher und außersprachlicher Vorgänge, ggf. mit Angabe der Dauer. Sie werden nur dann notiert, wenn sie relevant für die Interaktion sind. Neben Pausen, die in ihrer Länge geschätzt, aber auch zuweilen gemessen werden, gibt es die folgenden Klassen an NVV: “Verzögerungssignale” (Füller); “Lachen und Weinen”; “Rezeptionssignale” (Feedback-Äußerungen); “Nonverbale Handlungen und Ereignisse”, z. B. *Schniefen, Husten, Seufzen*. “Weitere phonetische Phänomene”, wie *Klicken und Schnalzen* oder *Ein- bzw. Ausatmen* nach bzw. zwischen intonatorischen Einheiten zählen zu den “Grenzsignalen”.

Die spontansprachlichen Korpora aus Kiel [9] haben eine umfangreichere Liste von NVV. Zu den Interjektionen zählen *au, ah, oh, aha, mhm* (“agreement”) und *mm* (“negation”). Die Klasse der nonverbalen Produktionen beinhaltet “Häsitationen” (Füller) mit Unterklassen wie *Husten, Lachen, Räuspern, Schlucken, Schmatzen* und einer Residualklasse “Geräusch” für alle weiteren nonverbalen artikulatorischen Geräusche. “Nicht-artikulatorische Geräusche” werden gesondert behandelt. Darüber hinaus wird “Atmung” als eigene Klasse betrachtet, aber nicht weiter unterteilt.

### 3 Methode

#### 3.1 Material

Wir verwenden Daten aus zwei Korpora, die zum einen aus spontanen Dialogen (GECO) und zum anderen aus professionell eingesprochenen Radionachrichten (DIRNDL) bestehen. Von beiden Korpora verwenden wir lediglich einen Auszug, der uns dennoch die Gegenüberstellung aufgabenfreier spontansprachlicher Dialoge mit geskripteten lesesprachlichen Monologen deutscher Muttersprachler ermöglicht. Aus GECO Version 1.1 [10] verwenden wir die ersten drei Minuten dreier Dialoge (sechs Sprecherinnen). Aus DIRNDL [11] extrahieren wir sechs Radionachrichten von drei Sprechern. Die Sprechdauer beträgt insgesamt ca. neun Minuten in den Dialogen und ca. vier Minuten in den Radionachrichten. Die Annotation erfolgt in Praat [12], die Suchanfrage mit *emuR* Version 1.1.1 [13].

#### 3.2 Annotationsschema

Die hier verwendete Mehr-Ebenen-Annotation enthält sechs Intervallebenen:

- Sprechaktivität (TURN)
- Pausen (PAUSE)
- Stille (SILENCE)
- Atmungsaktivität (RESP)
- Artikulatorische Aktivität (ARTIC)
- Residuale Phänomene (ELSE)

Auf der obersten Ebene (TURN) wurden Sprechphasen (Wert: *spk*) und Feedback-Äußerungen annotiert, die beide von ihrer konzeptuellen Definition und ihrem phonetischen Korrelat her nicht unproblematisch sind. Sprechphasen definieren wir als Abschnitte kontinuierlicher Sprachproduktion, die eben nicht nur bloße Feedback-Äußerungen sind und die auch nicht durch den Gesprächspartner unterbrochen werden, außer für eine kurze Feedback-Äußerung.

Als Feedback-Äußerungen (Wert: *fb*) definieren wir kurze lexikalische oder nicht-lexikalische Äußerungen, die prinzipiell verschiedene Funktionen annehmen können, wie z. B. Backchanneling, Zustimmung, Ablehnung oder Kommentare. Typische Beispiele für eine lexikalische Feedback-Äußerung sind *ja okay* oder *genau*, für eine nicht-lexikalische Feedback-Äußerung *mh* oder *hm*. In Übereinstimmung mit Schmidt [14] betrachten wir *hm* als einen neutralen Konsonanten bzw. Intonationsträger mit geschlossenem Mund. Feedback-Äußerungen diesen Typs bestehen häufig auch aus einem zweisilbigen *hm-hm*. Diese Partikeln enthalten keine lexikalische oder grammatische Information. Sprechphasen können vorhergehende oder nachfolgende Aktivität zeigen, die nicht Teil der eigentlichen Sprechphase sind und sich von Stille unterscheiden (wie bspw. Einatmungsgeräusche am Äußerungsbeginn). Links von einer Sprechphase annotieren wir daher antezedente Phasen (Wert: *ante*), rechts von einer Sprechphase postzedente Phasen (Wert: *post*). Die Antezedens-Phasen werden auf ein Maximum von 500 ms begrenzt, für die Postzedens-Phasen setzen wir den Wert auf die längste vorgekommene Länge (676 ms).

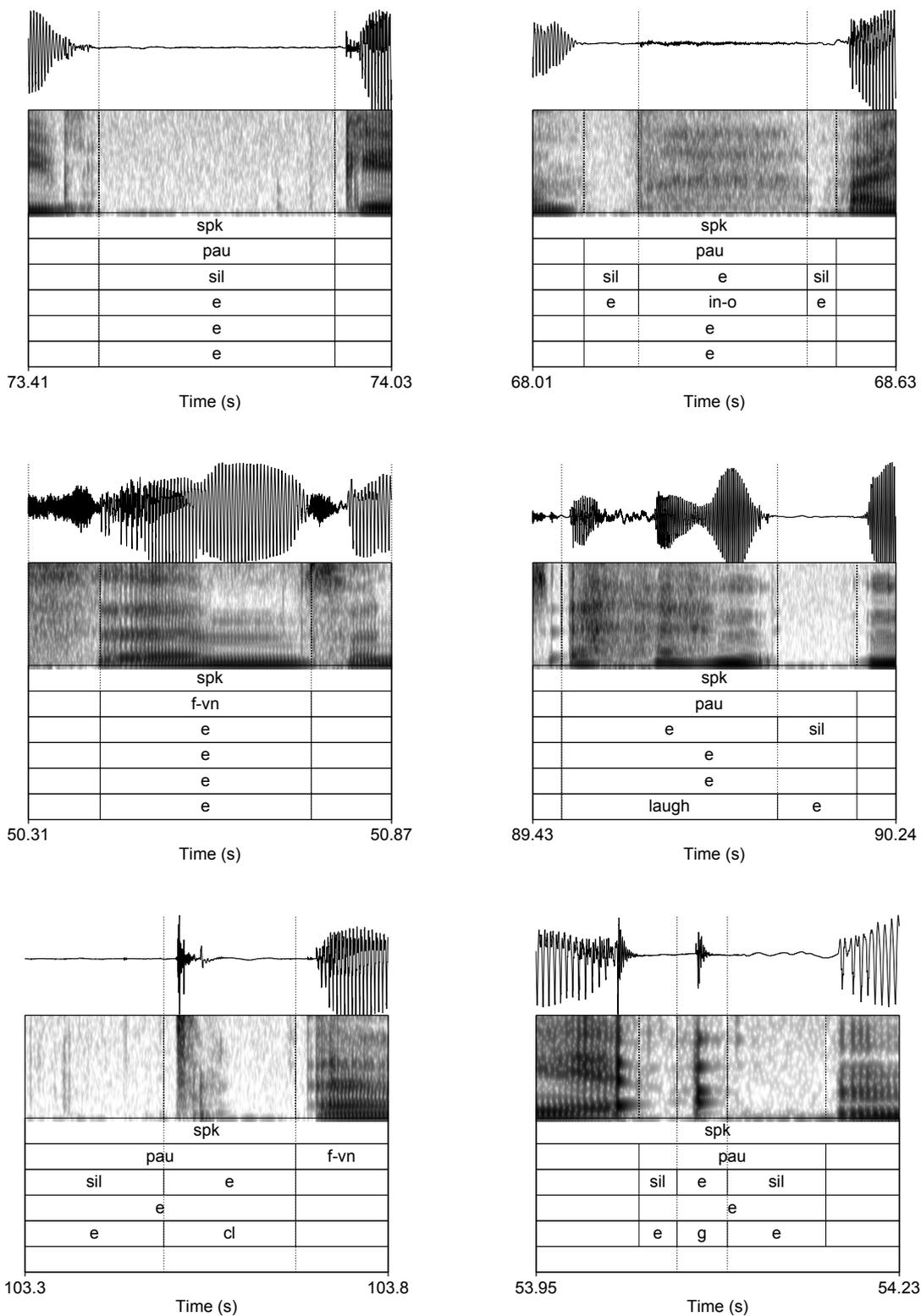
Auf der PAUSEN-Ebene markieren wir perzipierte Pausen innerhalb von *spk*-Phasen mit dem Wert *pau*. Akustisch lassen sich verschiedene mit *pau* annotierte Exponenten beobachten: zum einen ausschließlich Stille (s. Abbildung 1 links oben) und zum anderen Stille mit zusätzlichen NVV, wie z. B. Einatmungsgeräusche (s. Abbildung 1 rechts oben). Somit definieren wir *pau* als hörbare Unterbrechung des artikulatorischen Wortstroms innerhalb einer Sprechphase. Verschlussphasen von Plosiven (nach einer Pause nehmen wir den Richtwert von 50 ms für die Verschlussphase an) und prolongierte Verschlussphasen in Versprechern werden von dieser Definition ausgenommen. Weiterhin werden "gefüllte Pausen" auf dieser Ebene annotiert, die von uns als Teil von *spk* definiert werden. Typische Beispiele sind *äh* (Wert: *f-v*) als vokalischer Füller und *ähm* (Wert: *f-vn*) als vokalisch-nasale Sequenz. Beide Kategorien können vor dem Vokal Glottalplosivepenthesen zeigen. Andere Entitäten, die möglicherweise auch als Füller fungieren, sind Cluster aus Plosiven und Frikativen wie *pff* oder Sequenzen glottaler Plosive [15]. Diese erhalten den Wert *f-o* ('other fillers').

Auf der SILENCE-Ebene werden stille Phasen mit *sil* annotiert, die sich in unserem Fall auf die Sprechphasen beziehen und immer innerhalb den Grenzen einer Pause (*pau*) vorkommen. Stille Intervalle kommen natürlich auch außerhalb von Sprechphasen vor, werden dort aber nicht annotiert.

Respirationsereignisse und glottal-artikulatorische Aktivität bilden zwei voneinander getrennte Ebenen RESP und ARTIC, da diese auch gleichzeitig stattfinden können, wie z.B. Zungenschnalzlaut während der Einatmung [16]. Auf der RESP-Ebene wird Atmungsaktivität innerhalb der Intervalle *ante*, *spk* und *post* mit den Werten *in-o*, *in-n*, *ex-o* und *ex-n* für orale und nasale In- und Exhalation markiert. Auf der ARTIC-Ebene annotieren wir nicht-lexikalische artikulatorische Ereignisse wie Clicks (Wert: *cl*), die entweder als Einzelereignis oder als Reihe auftreten. Artikulatorisch nicht eindeutig bestimmbare, aber hörbare Fragmente werden mit *x* markiert.

Phänomene wie Lachen und vegetative Äußerungen wie Räuspern, Husten und Schlucken oder affektive Ausdrücke werden auf der ELSE-Ebene annotiert. Hinzu kommen noch akustisch auffällige nicht-verbale Ereignisse, die keine Vokalisierungen darstellen, wie z. B. Rascheln mit Papier oder Klopfen an der Tür. Lachen stellt einen Spezialfall dar. Oftmals ist es nicht Teil des Artikulationsflusses, wie die restliche Sprechphase, sondern es findet oftmals zusammen mit Einatmung statt, zuweilen kommen auch Sprechlacher während der Artikulation vor [16]. Lachen wird demgemäß innerhalb den Grenzen von *pau* auf der ELSE-Ebene annotiert.

Unklare Fälle können auf jeder Ebene mit *x* markiert werden. Zur besseren Auffindbarkeit von hierarchisch auf *pau* bezogenen Abschnitten werden die leeren Abschnitte auf den anderen Ebenen mit *e* für 'empty' gekennzeichnet (s. Abbildung 1).



**Abbildung 1** – Sechs Beispiele für Zeitsignale und Spektrogramme (5 kHz) von NVV mit allen sechs Annotationsebenen in der Reihenfolge TURN, PAUSE, SILENCE, RESP, ARTIC, ELSE (von oben nach unten). Alle Beispiele stammen von Sprecherin K im Korpus GECO [10]. Oben links: stille Pause; oben rechts: Atmungspause mit Einatmungsgeräusch; Mitte links: vokalisch-nasaler Füller *ähm*; Mitte rechts: Lachen (hier: Sprech-Lachen); unten links: Zungenlick; unten rechts: glottaler Reflex.

### 3.3 Herausforderungen für die Annotation

Die Abgrenzung der Sprechphasen von Feedback-Äußerungen ist nicht trivial, und Mehrdeutigkeiten wie in den folgenden Fällen können nicht immer klar desambiguiert werden. So kann eine Bestätigungsanfrage, die aus einem wiederholten Wort mit steigender Intonation besteht, sowohl als Sprechphase als auch als Feedback-Äußerung kategorisiert werden. Auch Lachen kann als Feedback interpretiert werden, wobei Feedback wiederum einen Turn initiieren kann [17].

Auf der Ebene PAUSE sind häufig Intervalle zwischen einer Feedback-Äußerung und einer Sprechphase zu beobachten, die mehrere Sekunden anhalten, oder aber kürzer als eine Sekunde sind. Diese Intervalle werden hier nicht als Pausen definiert, sondern könnten hypothetisch auch als Pausen zwischen Sprechphasen betrachtet werden.

Auf der Ebene SILENCE kann eine stille Phase, wie sie hier definiert wird, sichtbare Korrelate im Signal enthalten, die auch (aber nicht immer) hörbar sind, was für den Terminus *Stille* problematisch ist. Diese entstehen aufgrund von außen stammender nicht-vokalischer Geräusche (z.B. Türeenschlagen), aber auch aufgrund von unspezifischen Geräuschen, die von den Versuchspersonen stammen (z.B. Kratzgeräusche). Auf den ersten Blick widerspricht dieses Bild der Idee der Stille.

Auf der Ebene RESP werden Atmungsgeräusche annotiert, die auditiv wahrnehmbar sind. Sie werden in orale und nasale Ein- und Ausatemungsgeräusche unterteilt, was ohne weitere respiratorische Messungen nicht immer deutlich bestimmt werden kann.

Auf der Ebene ARTIC werden Clicks [18] (Wert: *cl*) annotiert. Diese können sowohl als Sequenzen wie auch als Einzelereignisse auftreten. Oft ist ihre Intensität eher gering, was die Annotationsentscheidung erschwert. Obwohl sie also im Spektrogramm gut sichtbar sind, sind sie nicht immer gut auditiv perzipierbar, wenn man den größeren Kontext einbezieht. In diesem Fall werden sie nicht annotiert. Weiterhin werden Instanzen glottaler Aktivität (glottale Plosive) annotiert (Wert: *g*). Diese können als Einzelinstanzen oder als Sequenzen vorkommen. Die Herausforderung besteht in der Trennung von Glottalisierung bzw. Laryngalisierung. Geräusche, die keiner eindeutigen phonetischen Kategorie zuordenbar sind, werden erhalten den Wert *else*.

## 4 Diskussion und Fazit

Tabelle 1 zeigt die Anzahl der jeweiligen Annotationswerte. Die Kategorie mit den häufigsten Items in beiden Stilen bilden Atemgeräusche, die in großer Mehrzahl als orale Einatmung produziert werden. Nur 16 % aller Pausen sind komplett still. In den Dialogen beträgt die relative Anzahl komplett stiller Pausen ca. ein Drittel (34 %). Einige Werte kommen nur in Dialogen und nicht in den Radionachrichten vor, z. B. Feedback-Äußerungen, Füller und Lachen. Hier wird nochmals deutlich, wie sehr sich die beiden Register gesprochener Sprache (spontan dialogisch vs. geskriptet monologisch) unterscheiden.

**Tabelle 1** – Anzahl der Annotationskategorien je Ebene für Dialoge und Radionachrichten.

		Dialoge	Nachrichten	Σ					
		Dialoge	Nachrichten	Σ	Dialoge	Nachrichten	Σ		
TURN	Feedback	96	0	96	RESP	Einatmung	133	56	189
	Sprechphase	81	6	87		Ausatmung	27	0	27
	Ante	58	1	59		Unklar	3	0	3
	Post	23	0	23	ARTIC	Click	32	16	48
	Post/ante	5	0	5		Glottal	16	0	16
PAU	Pause	144	67	211	Andere	13	3	16	
	Füller	29	0	29					
ELSE	Lachen	22	0	22					
	Andere	8	0	8					

Auch wenn eine geringe Anzahl von Pausen wirklich stille Pausen sind, so enthält doch die Mehrzahl der perzipierten Pausen NVV, mit Einatmungsgeräuschen als häufigster Kategorie. Es lassen sich etliche Funktionen dieser frikativähnlichen Laute sammeln, die aber noch einer viel gründlicheren und umfassenderen Untersuchung bedürfen. Entsprechend annotierte Korpora können dabei helfen.

Ein Blick auf andere Annotationsschemata lässt erkennen, dass wir weit von einem einheitlichen Schema entfernt sind. Dies betrifft zum einen das Inventar an NVV mit seinen Kategorien und Attributen, zum anderen aber auch die angenommene Anzahl von Ebenen. So lassen sich in den spontansprachlichen Dialogdaten wie zu erwarten etliche Lacher beobachten. Lachen ist ein gutes Beispiel dafür, dass es sich lohnt nicht nur eine, sondern mehrere Ebenen für die Annotation zu benutzen. Lacher können dabei als Sprechlacher vorkommen und damit auf der TURN-Ebene intervenieren, aber auch auf der RESP-Ebene.

Clicks und glottale Reflexe als NVV sind relativ neue Entdeckungen. Daher bleibt es zukünftigen Studien überlassen, deren Funktionen genauer zu bestimmen. Während glottale Reflexe tendenziell Ähnlichkeiten zu Füllern aufweisen, kommen Clicks als Gesten der Sprechvorbereitung in Frage.

## Literatur

- [1] WICHMANN, A.: *Speech corpora and spoken corpora*. In A. LÜDELING und M. KYTÖ (Hrsg.), *Corpus Linguistics*, Handbücher zur Sprach- und Kommunikationswissenschaft, S. 187–207. Walter de Gruyter, Berlin, 2008.
- [2] TROUVAIN, J. und K. P. TRUONG: *Comparing non-verbal vocalisations in conversational speech corpora*. In *4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals*, S. 36–39. 2012.
- [3] RHODES, R.: *Aural images*. In L. HINTON, J. NICHOLS, und J. OHALA (Hrsg.), *Sound Symbolism*, S. 276 – 292. Cambridge University Press, Cambridge, 2006.
- [4] BELZ, M., S. SAUER, A. LÜDELING, und C. MOOSHAMMER: *Fluently disfluent? Pauses and repairs of advanced learners and native speakers of German*. *International Journal of Learner Corpus Research*, 3(2), S. 118–148, 2017. doi:10.1075/ijlcr.3.2.02bel.
- [5] BORTFELD, H., S. D. LEON, J. E. BLOOM, M. F. SCHOBBER, und S. E. BRENNAN: *Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender*. *Language and Speech*, 44(2), S. 123–147, 2001. doi:10.1177/00238309010440020101.
- [6] TROUVAIN, J. und K. TRUONG: *Comparing non-verbal vocalisations in conversational speech corpora*. In *Proc. 4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, S. 36 – 39. Istanbul, 2012.
- [7] VAN ENGEN, K. J., M. BAESE-BERK, R. E. BAKER, A. CHOI, M. KIM, und A. R. BRADLOW: *The wildcat corpus of native- and foreign-accented english: Communicative efficiency across conversational dyads with varying language alignment profiles*. *Language and Speech*, 53(4), S. 510–540, 2010. doi:10.1177/0023830910372495.
- [8] SELTING, M., P. AUER, D. BARTH-WEINGARTEN, J. BERGMANN, P. BERGMANN, K. BIRKNER, E. COUPER-KUHLEN, A. DEPPERMAN, P. GILLES, S. GÜNTNER, M. HARTUNG, F. KERN, C. MERTZLUFFT, C. MEYER, M. MOREK, F. OBERZAUCHER, J. PETERS, U. QUASTHOFF, W. SCHÜTTE, A. STUKENBROCK, und S. UHMANN:

*Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)*. *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion*, 10, S. 353–402, 2009.

- [9] KOHLER, K. J., M. PÄTZOLD, und A. P. SIMPSON: *From scenario to segment - the controlled elicitation, transcription, segmentation and labelling of spontaneous speech*, Bd. 29 d. Reihe AIPUK. Kiel, 1995.
- [10] SCHWEITZER, A. und N. LEWANDOWSKI: *Convergence of articulation rate in spontaneous speech*. In *Proc. Interspeech*, S. 525 – 529. Lyon, 2013.
- [11] ECKART, K., A. RIESTER, und K. SCHWEITZER: *A discourse information radio news database for linguistic analysis*. In C. CHIARCOS, S. NORDHOFF, und S. HELLMANN (Hrsg.), *Linked Data in Linguistics*, S. 65 – 75. Springer, Berlin and Heidelberg, 2012.
- [12] BOERSMA, P.: *Praat, a system for doing phonetics by computer*. *Glott International*, 5(9), S. 341–345, 2001.
- [13] WINKELMANN, R., K. JAENSCH, S. CASSIDY, und J. HARRINGTON: *emur: Main package of the emu speech database management system*. 2018.
- [14] SCHMIDT, J. E.: *Bausteine der Intonation? Germanistische Linguistik*, 157-158, S. 9–32, 2001.
- [15] BELZ, M.: *Glottal filled pauses in German*. In R. EKLUND und R. L. ROSE (Hrsg.), *Proceedings of DiSS 2017, TMH-QPSR*, S. 5–8. 2017.
- [16] TROUVAIN, J.: *Laughing, breathing, clicking – the prosody of nonverbal vocalisations*. In *Proc. Speech Prosody*, S. 598 – 602. Dublin, 2014.
- [17] DRUMMOND, K. und R. HOPPER: *Back channels revisited: Acknowledgment tokens and speakership incipiency*. *Research on Language & Social Interaction*, 26(2), S. 157–177, 1993. doi:10.1207/s15327973rlsi2602\_3.
- [18] TROUVAIN, J. und Z. MALISZ: *Inter-speech clicks in an Interspeech keynote*. In *Inter-speech*, S. 1397–1401. 2016.