

Reliabilität zwischen Annotatoren in DIMA –

Ein Vergleich der Annotation von Experten und einer Trainingsgruppe

Phonetik und Phonologie im deutschsprachigen Raum 13, Berlin, 28.-29. September 2017

Frank Kügler^{1,2}, Stefan Baumann¹, Antje Schweitzer³ & Petra Wagner⁴

¹IfL Phonetik, Universität zu Köln, ²Goethe Universität Frankfurt, ³Universität Stuttgart, ⁴Universität Bielefeld
 {fkuegler,stefan.baumann}@uni-koeln.de, antje.schweitzer@ims.uni-stuttgart.de, petra.wagner@uni-bielefeld.de



Universität Bielefeld

Hintergrund und Motivation

DIMA – „Deutsche Intonation: Modellierung und Annotation“ [1] – ist ein Konsenssystem zur Annotation der deutschen Intonation.

DIMA unterscheidet drei Ebenen :

(1) Phrasen (2) Töne (3) Prominenz

Besondere Eigenschaften:

- DIMA-Annotation erfordert keine abschließende phonologische Interpretation der Daten.
- DIMA trennt das komplexe Intonationssignal und lenkt die Perzeption der Intonation unabhängig voneinander auf die drei Ebenen.

Symbole für die Annotation

Ebene	Symbole
Phrase	% -
Diakritika	! ^ & r e ?
Ton	H* L* H L
Diakritika	! ^ < > ?
Prominenz	1 2 3
Diakritika	?
Kommentar	z.B. alternative Annotation

Ziel der Studie

DIMA-Annahme: das System ist leicht trainierbar.

Hypothese: Eine Trainingsgruppe kann das DIMA System leicht erlernen.

ZIEL: Vergleich der Reliabilität der Annotation zwischen Experten und einer Trainingsgruppe.

Methode

- Zwei Gruppen: (i) Experten (die Autoren) vs. (ii) Trainingsgruppe
- Trainingsgruppe: 8 Studenten der Uni Köln (5w/3m)
 - DIMA-Training in 4 Sitzungen zu 1h mit Feedback
- Sprachdaten – Korpus SFB 732 *silver standard collection* [3]:
 - spontansprachlicher Ausschnitt, politisches Interview (2 min)
 - 441 Wörter
- Annotation nach DIMA-Richtlinien, Version 3.05 [4]
- Fleiss-Kappa [κ] Berechnung der An- bzw. Abwesenheit von Labeln [5]

Ergebnisse

Korpus	2016 [κ]	Experten [κ]	Trainees [κ]
Interview (gesamt)	0,73	0,675	0,693

substantial

Tab. 1: Gesamtdaten; Einschluss aller drei Ebenen.

Skala der Übereinstimmung [5]:

0,41 – 0,6: “moderate agreement” 0,61 – 0,8: “substantial agreement”

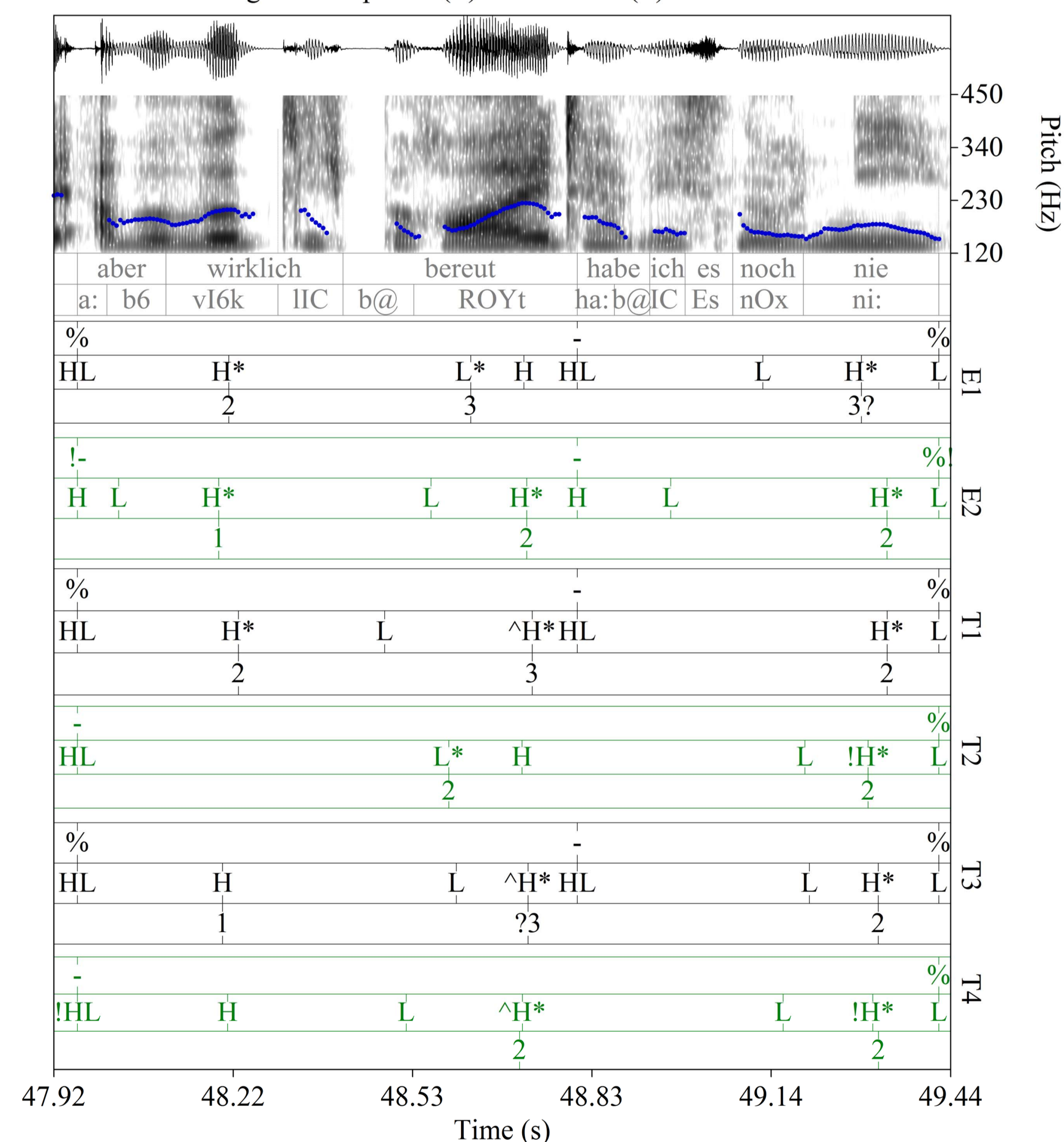
Gruppe	Phrase	agreement	Ton	agreement	Prominenz	agreement
2016 [κ]	0,621	substantial	0,506	moderate	0,627	substantial
Experten [κ]	0,697	substantial	0,45	moderate	0,634	substantial
Trainees [κ]	0,657	substantial	0,448	moderate	0,54	moderate

Tab. 2: Übereinstimmung per Ebene.

Gruppe	Phrase*Ton	agreement	Prominenz*Ton	agreement
2016 [κ]	0,784	substantial	0,678	substantial
Experten [κ]	0,78	substantial	0,695	substantial
Trainees [κ]	0,77	substantial	0,64	substantial

Tab. 3: Übereinstimmung von Phrasen und Tönen an Phrasengrenzen sowie von Prominenz und Tönen an Prominenzen.

Vergleich: Experten (E) und Trainee (T) Annotation



Schlussfolgerung

- **Insgesamt substantielle Übereinstimmung beider Gruppen!**
- Übereinstimmung der Trainees entspricht Übereinstimmung der Experten.
- Moderate Übereinstimmung auf der Tonebene ist vergleichbar zu geringerer Übereinstimmung bei der GToBI Annotation von Akzenttypen (vgl. [6]).
- **DIMA IST relativ leicht erlernbar!**
- Trainingsmaterial auf DIMA-Webseite <http://dima.uni-koeln.de/>

Danksagung:

Ein Dank gilt der gesamten DIMA-Gruppe, die zu den Richtlinien 3.05 beigetragen hat. Dank auch an die Trainees dieser Studie.

References

- [1] Kügler, F., Smolibocki, B., Arnold, D., Braun, B., Baumann, S., Grice, M., Jannedy, S., Michalsky, J., Niebuhr, O., Peters, J., Ritter, S. B., Röhr, C. T., Schweitzer, A., Schweitzer, K. Wagner, P. 2015. DIMA - Annotation guidelines for German intonation. In: Proc 18th ICPhS (paper 317: 1-5). Glasgow.
- [2] Kügler, F., Baumann, S., Schweitzer, A. & Wagner, P. 2016. DIMA – Erweiterte Richtlinien zur Annotation deutscher Intonation. Vortrag, P&P-12, München16.
- [3] Eckart, K. & Gärtner, M., Creating Silver Standard Annotations for a Corpus of Non-Standard Data. In: S. Dipper et al. (eds), Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016) volume 16 of BLA: Bochumer Linguistische Arbeitsberichte pp. 90-96 Bochum, Germany.
- [4] Kügler, F. & Baumann, S. 2017. Annotationsrichtlinien DIMA. (<http://dima.uni-koeln.de/>).
- [5] Fleiss, J. L. 1971. Measuring nominal scale agreement among many raters. Psychological Bulletin, 76, 378–382.
- [6] Reyelt, M., Grice, M., Benz Müller, R., Mayer, J., Batliner, A. 1996. Prosodische Etikettierung des Deutschen mit ToBI. VM-Report 154, Saarbrücken.