Tatort Sprachverständlichkeit – Wie schlimm ist es wirklich?

Rainer Huber, Hannah Baumgartner, Tobias Bruns, Julia Thomas, Jens-E. Appell, Jan Rennies

Fraunhofer IDMT, Hör-, Sprach- und Audiotechnologie, Oldenburg

Abbildung 1: Szene aus dem Wiener Tatort
(Foto: ORF/Petro Domenigg)

Einleitung

Der "Tatort" ist eine der meistgesehenen Sendungen der ARD. In Zuschauerkommentaren wird häufig eine schlechte Sprachverständlichkeit kritisiert [1, 2]. Dieser Vorwurf soll durch eine instrumentelle Methode zur Messung der Höranstrengung [3] überprüft werden. Dabei werden verschiedene Tatortproduktionen sowohl untereinander als auch mit anderen TV-Produktionen verglichen.

Methode

Die in [3] vorgestellte Methode zur Vorhersage der Höranstrengung beruht auf der Verwendung eines tiefen neuronalen Netzes eines automatischen Spracherkennungssystems zur Berechnung von Phonem-posterior-Wahrscheinlichkeiten („Posteriorgramme“). Störungen der Sprache, z.B. durch Hintergrundgeräusche, erhöhen die Erkennungsunsicherheit des Erkennungssystems, was sich durch „Verschmierungen“ der Posteriorgramme äußert. Der Verschmierungsgrad wird durch das mathematische M-Measure nach Hermansky et al. [4] quantifiziert. Dieses Maß dient als Prädiktor für die Höranstrengung. Seine Eignung zur Vorhersage der Höranstrengung im TV-Bereich konnte in [3] gezeigt werden. Abbildung 2 illustriert die Vorhersagegüte anhand eines Scatterplots mit über 400 Bewertungen verschiedener Ausschnitte aus deutschen Fernsehsendungen. Die hier verwendete Implementation der Methode schätzt die Höranstrengung alle 30ms für jeweils rund 1s lange Signalabschnitte, die laut automatischer Sprachaktivitätserkennung [5] Sprache enthalten.

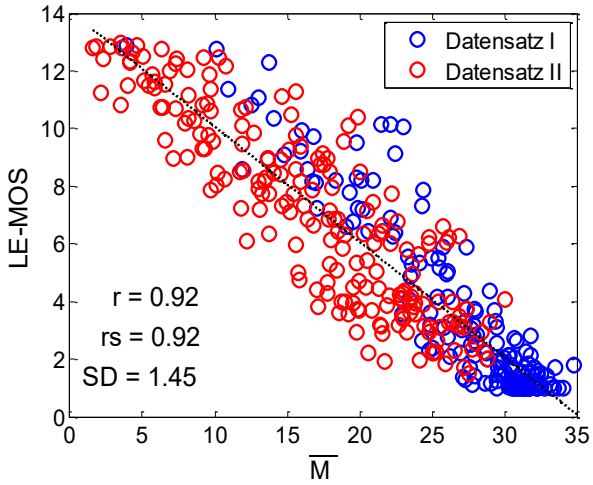

Abbildung 2: Gegenüberstellung von über die Probanden gemittelten subjektiven Höranstrengungsbewertungen (“Listening Effort Mean Opinion Scores” – LE-MOS) und entsprechenden Werten des Prädiktors M für die Datensätze zweier Hörexperimente mit TV-Audiosignalen. r: Korrelationskoeffizient nach Pearson. rs: Rang-Korrelationskoeffizient nach Spearman (aus [3]).

Datensätze

Zur Analyse der Höranstrengung mit der beschriebenen instrumentellen Methode wurden die folgenden Sendungen aus der ARD-Mediathek heruntergeladen (in Klammern die Darsteller/innen sowie Anzahl der jeweiligen Sendungen):

- Tagesschau (N = 10)
- Mankells Wallander (N = 10)
- Tatort Berlin (Becker & Waschke; N = 6)
- Tatort Frankfurt (Sawatzki & Schüttauf; N = 5)
- Tatort Hamburg (Möhring & Weisz; N = 4)
- Tatort Köln (Behrendt & Bär; N = 10)
- Tatort München (Nemec & Wachtveitl; N = 10)
- Tatort Münster (Prahl & Liefers; N = 8)
- Polizeiruf Rostock (Hübner & Sarnau; N = 6)
- Tatort Stuttgart (Müller & Klare; N = 7)
- Tatort Weimar (Tschirner & Ulmen; N = 8)
- Tatort Wien (Krassnitzer & Neuhauser; N = 4)

Ergebnisse

Die Tonspuren der o.g. Fernsehsendungen wurden durch die instrumentelle Methode hinsichtlich ihrer Höranstrengung analysiert. Die Ergebnisse der Analyse sind in Form von exemplarischen Histogrammen (Häufigkeit über Höranstrengungswert) in Abbildung 3 sowie in Form eines Boxplots in Abbildung 4 dargestellt. Die Höranstrengung ist dabei in *Effort Scaling Categorical Units* (ESCU) angegeben und reicht von 1 – *mühelos* bis 13 – *extrem anstrengend*.

Zwischenkategorien sind: 3 – *sehr wenig anstrengend,* 5 – *wenig anstrengend*, 7 – *mittelgradig anstrengend,* 9 – *deutlich anstrengend*, 11 – *sehr anstrengend.* In jedem Histogramm sind darüber hinaus der Median sowie das 75. Perzentil der Häufigkeitsverteilung angegeben. Der Parameter N im Titel der Histogramme gibt die Zahl der Sendungen, die jeweils in die Analyse eingeflossen sind, an.

Die Histogramme der Krimisendungen sind untereinander recht ähnlich (Abbildung 3 zeigt die beiden Extrema mit dem kleinsten bzw. größten 75. Perzentil der Höranstrengung) und unterscheiden sich deutlich von dem der Tagesschau. Die Mediane der Verteilungen der Krimisendungen liegen im Bereich 4,2 bis 5 ESCU; die 75. Perzentile überstreichen einen Bereich von 5,9 bis 7,3 ESCU. Dagegen fallen Median und 75. Perzentil der Tagesschau mit 1,8 bzw. 2,4 ESCU deutlich kleiner aus.

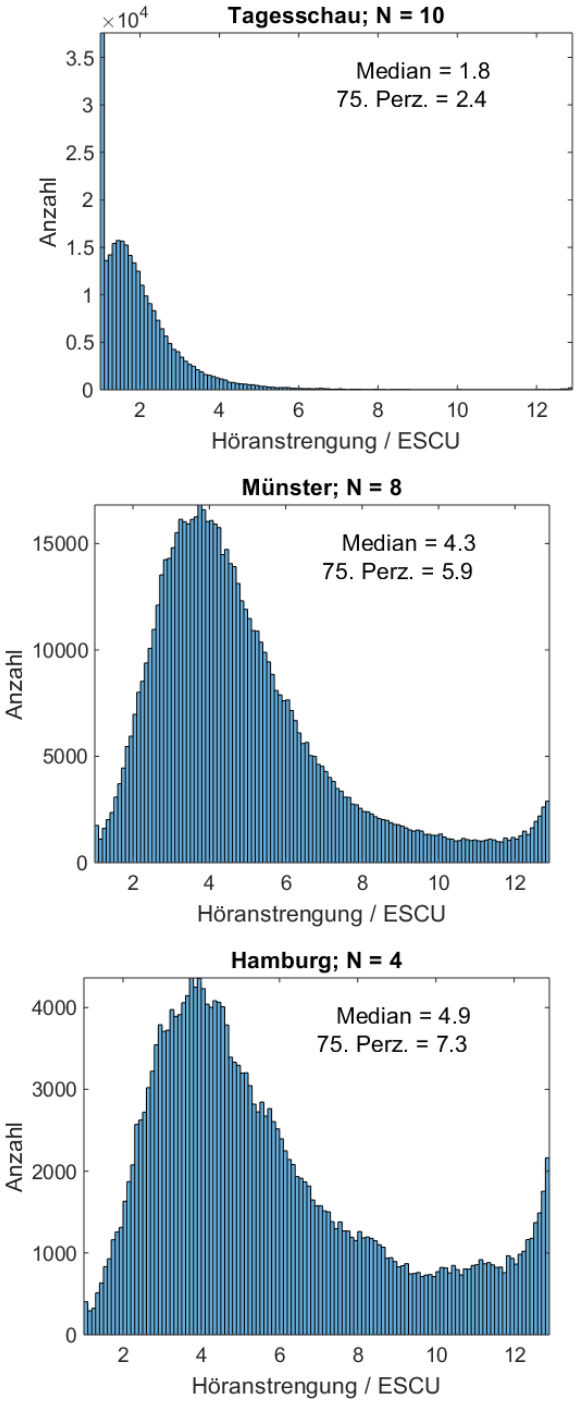

Abbildung 3: Exemplarische Histogramme der berechneten Höranstrengungswerte für die Tagesschau (oben) sowie zweier Tatortproduktionen. „N“ bezeichnet die Zahl der jeweiligen Filme/Sendungen. Die Höranstrengung, angegeben in *Effort Scaling Categorical Units* (ESCU) reicht von 1 – *mühelos* bis 13 – *extrem anstrengend*.

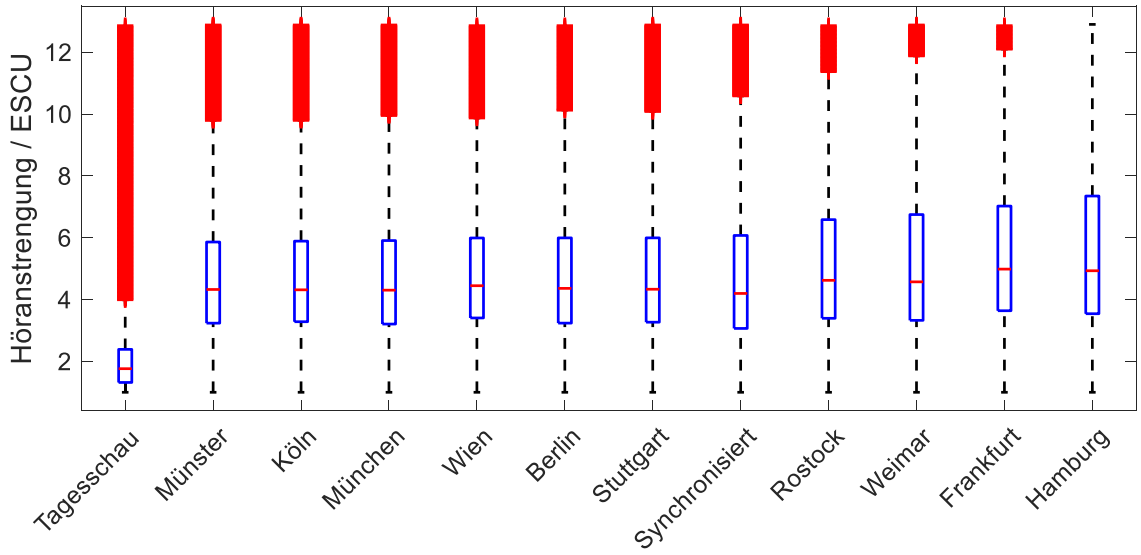

Abbildung 4: Boxplot der berechneten Höranstrengungswerte für die verschiedenen Sendungen, sortiert nach aufsteigendem 75. Perzentilwerten. Jede Box zeigt den Median (roter, horizontaler Strich innerhalb der Box); Unter- und Oberkante der Box werden bestimmt durch das untere bzw. obere Quartil. Die Enden der gestrichelten Linien geben die minimalen bzw. maximalen Werte an. Rote Rechtecke symbolisieren Ausreißer. Die Höranstrengung, angegeben in *Effort Scaling Categorical Units* (ESCU) reicht von 1 – *mühelos* bis 13 – *extrem anstrengend*.

Diskussion

Die Ergebnisse der instrumentellen Höranstrengungsbewertung zeigen einen deutlichen Unterschied zwischen den Höranstrengungswerten der Krimisendungen einerseits und der Tagesschau andererseits. Dies war zu erwarten, da die Sprache der Tagesschau einerseits überwiegend nicht mit Hintergrundgeräuschen oder -musik überlagert ist und andererseits die Sprache wohlartikuliert ist, im Gegensatz zu den Fernsehfilmen, zumindest, was die Hintergrundgeräusche und -musik betrifft. Aber auch undeutliche Aussprachen der Darsteller können bei den Fernsehfilmen auftreten und möglicherweise einen Teil der Differenzen zwischen den Höranstrengungs-Verteilungen verschiedener Fernsehfilme erklären. Die andere mögliche Ursache sind systematisch unterschiedliche Sprache- zu Hintergrund-Mischungsverhältnisse zwischen den verschiedenen Produktionsstandorten. Während eine korrekte Berücksichtigung des Einflusses variierender Sprache- zu Hintergrundverhältnisse bei der Modellierung der Höranstrengung von Fernsehsendungen gezeigt werden konnte [3], steht ein entsprechender Nachweis für den Einfluss undeutlicher Aussprache noch aus. Das vorliegende Ergebnis einer im Vergleich zu den übrigen Fernsehfilmen erhöhten Höranstrengung des Tatorts Weimar mit den wegen ihrer z.T. undeutlichen Aussprache kritisierten Darsteller Ulmen/Tschirner [2] darf als Hinweis auf eine Sensitivität des Höranstrengungsmodells gegenüber undeutlicher Aussprache gedeutet werden.

Die Platzierung der synchronisierten Krimireihe „Mankells Wallander“ im unteren Bereich der Höranstrengung (d.h. mit rel. niedrigen Höranstrengungswerten) ist plausibel, da die Sprachqualität von im Studio aufgenommener Sprache i.A. besser ist als solche, die draußen am Aufnahmeset aufgenommen wurde. Die Stärke des Effekts war allerdings noch größer als hier beobachtet erwartet worden.

Überhaupt sind die beobachteten Unterschiede zwischen den Verteilungen bzw. Histogrammen der Höranstrengungswerte der verschiedenen Fernsehfilme relativ gering; geringer vielleicht, als angesichts der berichteten empfundenen Sprachverständlichkeitsunterschiede zwischen den verschiedenen Tatort-Standorten bzw. Darstellern erwartbar gewesen wäre. Dies mag daran liegen, dass bei der subjektiven Beurteilung der Sprachverständlichkeit einer Fernsehsendung wenige nicht oder schlecht verstandene Sätze überproportional zum Gesamteindruck beitragen können. Diese vielleicht wenigen, aber u.U. entscheidenden nicht verstandenen Sätze können in der Statistik „untergehen“ bzw. dort nur schwach repräsentiert sein. Ein anderer möglicher Grund für die relativ geringen Unterschiede zwischen den Höranstrengungsstatistiken ist die bereits oben in Frage gestellte ausreichende Sensitivität des Modells gegenüber undeutlicher Aussprache. Um dies zu klären, sind Evaluationen mit subjektiv bewerteten Sprachbeispielen mit variierender Artikulationsgüte nötig und geplant.

Zusammenfassung und Fazit

Das in [3] vorgestellte Tool zur instrumentellen Bewertung der Höranstrengung von Fernsehsendungen wurde auf eine Reihe von Tatort-Fernsehfilmen angewandt, um die in der Öffentlichkeit berichteten, teilweise kritikwürdigen Sprachverständlichkeiten dieser Sendungen durch Messwerte objektiv zu überprüfen. Dabei wurden in der Tat Unterschiede zwischen verschiedenen Ermittlerteams festgestellt, die tendenziell mit den berichteten subjektiven Eindrücken übereinstimmen; diese Unterschiede in den Statistiken der objektiven Metrik fielen jedoch relativ gering aus. Bei sechs der zehn überprüften Tatort-Ermittlerteams war die Höranstrengung laut Statistik sogar etwa gleich wie die einer synchronisierten Krimiserie mit vermeintlich besserer Sprachverständlichkeit. Eine pauschale schlechte Sprachverständlichkeit aller Tatort-Produktionen konnte damit nicht

bestätigt werden. Auch wenn der Nachweis einer angemessenen Sensitivität des Tools gegenüber undeutlicher Aussprache („nuscheln“) noch aussteht, so erscheint es bereits jetzt als nützliches Hilfsmittel zur objektiven Evaluation der Höranstrengung von Fernsehsendungen.

Literatur

[1] Der Tagesspiegel: “Das große Nuscheln”, https://www.tagesspiegel.de/gesellschaft/medien/schlechte-tonqualitaet-im-tv-das-grosse-nuscheln/13457678.html

[2] Focus online: “’Tatort’ - Zuschauer klagen über ‘miserablen’ Ton: So erklärt die ARD das Dauerproblem”, https://www.focus.de/kultur/kino_tv/tatort-zuschauer-klagen-ueber-miserablen-ton-so-erklaert-die-ard-das-dauerproblem_id_6567088.html

[3] Huber, R., Baumgartner, H., Rollwage, C., Goetze, S., Rennies, J.: Erfassung der Höranstrengung fertiger TV-Mischungen, In: Fortschritte der Akustik, DAGA 2019, DEGA, Berlin.

[4] Hermansky, H., Variani, E., Peddinti, V.: Mean temporal distance: Predicting ASR error from temporal properties of speech signal. In Proceedings ICASSP 2013, 38th IEEE Int. Conf. Acoust. Speech Signal Process. doi: 10.1109/ICASSP.2013.6639105.

[5] Moritz, N., Drefs, J., Baumgartner, H., Rennies, J.: Sprachaktivitätserkennung basierend auf Deep Neural Networks für Anwendung in Film und Fernsehen. In Fortschritte der Akustik. DAGA 2016, S.960-963; DEGA, Berlin.