

PERZEPTION VON SPRECHGESCHWINDIGKEIT UND DER (NICHT NACHGEWIESENE) EINFLUSS VON SURPRISAL

Frank Zimmerer¹, Bistra Andreeva¹, Bernd Möbius¹, Zofia Malisz², Emmanuel Ferragne³, François Pellegrino⁴, Erika Brandt¹

¹*Sprachwissenschaft und Sprachtechnologie, Universität des Saarlandes, Saarbrücken*

²*Department of Speech, Music and Hearing, KTH Stockholm*

³*CLILLAC-ARP, EA 3967, Université Paris Diderot*

⁴*Laboratoire Dynamique du langage, UMR 5596 CNRS Université Lyon 2
zimmerer@coli.uni-saarland.de*

Kurzfassung: In zwei Perzeptionsexperimenten wurde die Perzeption von Sprechgeschwindigkeit untersucht. Ein Faktor, der dabei besonders im Zentrum des Interesses steht, ist Surprisal, ein informationstheoretisches Maß für die Vorhersagbarkeit einer linguistischen Einheit im Kontext. Zusammengenommen legen die Ergebnisse der Experimente den Schluss nahe, dass Surprisal keinen signifikanten Einfluss auf die Wahrnehmung von Sprechgeschwindigkeit ausübt.

1 Motivation und Hintergrund

Die Wahrnehmung der (intendierten) Sprechgeschwindigkeit ist abhängig von vielen Faktoren, die über akustische Signaleigenschaften hinausgehen [z.B. 1, 2, 3, 4]. In einer Perzeptionsstudie konnten Dellwo et al. [4] zeigen, dass die *intendierte* Sprechgeschwindigkeit von Hörern wahrgenommen wird. Dieser quasi-kategoriale Effekt trat auf, obwohl Stimuli verschiedener intendierte Geschwindigkeiten bezüglich ihrer *gemessenen* Sprechgeschwindigkeit Überlappungen zeigten. Zudem konnte der Effekt für Stimuli verschiedener Sprachen nachgewiesen werden. Koreman [2] präsentierte Evidenz, dass nicht alleine die Anzahl der realisierten Segmente pro Zeit die Perzeption von Sprechgeschwindigkeit beeinflusste. Je nachdem, wie stark Personen in einem informellen Redestil die Segmente reduzierten, wurden unterschiedliche Sprechgeschwindigkeiten wahrgenommen. In einer späteren Studie replizierte Koreman [3] diese Ergebnisse.

In der vorliegenden Studie soll der Einfluss von Surprisal (oder Vorhersagbarkeit) auf die Perzeption von Sprechgeschwindigkeit untersucht werden. Surprisal, ein informationstheoretisches Maß für die Wahrscheinlichkeit, dass eine linguistische Einheit (Wort, Silbe, Laut) in einem bestimmten Kontext auftritt, wurde als ein Faktor identifiziert, der die Produktion von Sprache beeinflusst [z.B. 5, 6, 7, 8, 9]. Eine Annahme, die in diesen Studien im Mittelpunkt steht, ist es, dass Sprecher ihre Produktion verändern, um die Informationsdichte in einem gewissen (optimalen) Bereich zu halten.¹

Auch die Perzeption ist zumindest implizit in Produktionsstudien mit einbezogen. Was den genauen Einfluss von Surprisal auf die Perzeption betrifft, sind bislang weniger Studien durchgeführt worden, allerdings gibt es Resultate die solche einen Einfluss nahe legen. Dethlefs et al. [10] konnten in einem Rating Experiment zeigen, dass Hörer sensitiv für Surprisal (beziehungsweise *information density*) in der Perzeption sind. Allerdings blieben in der Studie einige Fragen

¹Wir unterscheiden hier nicht zwischen Informationsdichte, Surprisal und Vorhersagbarkeit, da diese voneinander abgeleitet werden können.

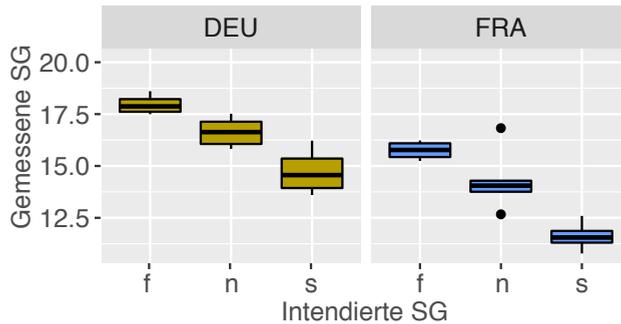


Abbildung 1 – Tatsächlich gemessene (in Anzahl der Laute / Sekunde) und intendierte (schnell (f), normal (n) und langsam (s)) Sprechgeschwindigkeit (SG) der Stimuli in Experiment 1.

ungeklärt. Eine dieser Fragen beschäftigt sich mit der Wahrnehmung von Sprechgeschwindigkeit. Um den Einfluss von Surprisal auf die Perception von Sprechgeschwindigkeit genauer zu beleuchten, wurden 2 Experimente durchgeführt, die im nächsten Abschnitt genauer dargestellt werden.

2 Experimente

Die beiden Perzeptionsstudien (ein Diskriminationsexperiment und ein Skalierungsexperiment) verwenden Stimuli des BonnTempo Corpus (BTC) [11]. Aus den Sätzen, die im BTC gelesen wurden, wurden (Audio-)Intervalle mit einer Dauer zwischen 1 s und 1,5 s herausgeschnitten. Diese Stimuli beinhalten keine Pausen und überschreiten keine Satzgrenzen. Anfang und Ende der Intervalle wurden bearbeitet, um Klicks und andere Störgeräusche zu vermeiden.

Die Stimuli variieren in der gemessenen Sprechgeschwindigkeit. Außerdem weisen sie auch unterschiedliche intendierte Sprechgeschwindigkeiten auf. Diese Variation führt zu teils überlappenden gemessenen Sprechgeschwindigkeiten in unterschiedlichen intendierten Sprechgeschwindigkeiten (vgl. Abbildung 1 für Experiment 1).

2.1 Experiment 1

Das Diskriminationsexperiment wurde durchgeführt, um zu untersuchen, ob Hörer in der Lage sind, die intendierte Sprechgeschwindigkeit zu erkennen, selbst wenn die gemessene Sprechgeschwindigkeit dazu im Widerspruch steht. Ein weiterer zentraler Aspekt des Experimentes war es herauszufinden, ob Hörer auf die gemessene Sprechgeschwindigkeit zurückgreifen, wenn die intendierte Sprechgeschwindigkeit sich nicht unterscheidet.

2.1.1 Materialien

Für die Diskriminationsstudie wurden deutsche und französische Stimuli von 4 Sprechern (2 Muttersprachler des Deutschen (2 Männer), 2 des Französischen (1 Frau, 1 Mann)) in drei Stufen des intendierten Sprechtempos (schnell, normal, langsam) verwendet. Von allen Sprechern wurden je 7 Stimuli verwendet (2 schnell (f), 2 langsam (s), 3 normal (n)). Jeweils 2 Stimuli wurden (innerhalb einer Sprache) gepaart, sodass die intendierte Sprechrate entweder gleich schnell war oder sich um eine Stufe unterschied. Es gab keine Paarung von schnellen mit langsamen Stimuli.

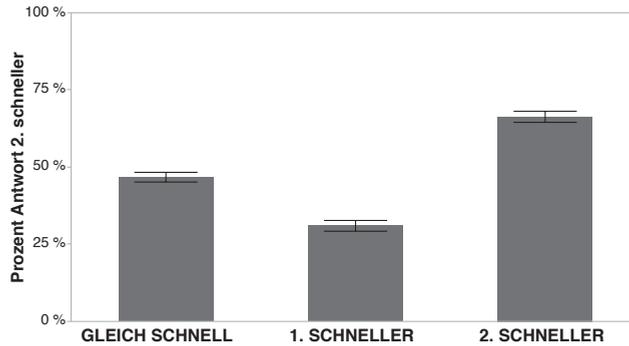


Abbildung 2 – Experiment 1: Prozentualer Anteil der Wahrnehmung “zweiter Stimulus ist schneller” für Stimuli mit gleicher und verschiedener (hier: erster Stimulus intendiert schneller vs. zweiter Stimulus intendiert schneller) intendierter Sprechgeschwindigkeit.

2.1.2 Versuchspersonen

Insgesamt wurden 11 deutsche Muttersprachler getestet (3 männlich, 8 weiblich - Durchschnittsalter: 26,7). Alle waren Studenten an der Universität des Saarlandes in Saarbrücken und wurden für ihre Teilnahme monetär entschädigt. Ihre Kenntnisse der französischen Sprache waren unterschiedlich.

2.1.3 Methode

Das Experiment wurde mit der Software E-Prime² durchgeführt. Insgesamt wurden den Probanden 220 Stimuluspaare vorgespielt. Ihre Aufgabe war es, zu entscheiden, welches Intervall des jeweiligen Paares schneller war. Diese Entscheidung konnten die Probanden per Fingerdruck auf eine Antwortbox abgeben. Vor jedem der beiden Intervalle wurde ein Signalton abgespielt, damit die Probanden erkennen konnten, wann ein neues Intervall begann. Die Pause zwischen zwei Stimuli eines Paares betrug 500 ms. Für ihre Antwort hatten die Probanden 4 s Zeit. Einige wenige (3 von 2420 Antworten, 0,1%) waren langsamer als 4 s und wurden von der weiteren Analyse ausgeschlossen.

2.1.4 Ergebnisse

Wenn die Intervalle hinsichtlich ihrer intendierten Sprechgeschwindigkeit gleich schnell waren (d.h. n vs. n , f vs. f , s vs. s), antworteten die Probanden in 46,7% der Fälle, dass das zweite Intervall schneller war. War das zweite Intervall intendiert schneller, dann wurde es zu 66,3% entsprechend wahrgenommen. Für Intervallkombinationen, in denen das erste Intervall das intendiert schnellere war, empfanden die Probanden in 31% der Fälle das zweite Intervall als schneller (Abbildung 2). Ein logistisches Regressionsmodell mit intendierter Sprechgeschwindigkeit (INTSG (f , n , s)), gemessener Sprechgeschwindigkeit (LABSG), logarithmierter Reaktionszeit sowie Teilnehmer (als Zufallsfaktor) zeigte, dass die Entscheidung, welches Intervall schneller ist, im Fall von gleichen intendierten Sprechgeschwindigkeiten zufällig ist. In allen Fällen, in denen sich die Sprechgeschwindigkeit der Intervalle unterschied, waren die Probanden signifikant erfolgreicher. Die Reaktionszeit reflektierte nicht, ob es sich um eine schwierige Entscheidung (d.h. gleich schnelle Items) handelte, oder ob die Intervalle aus verschiedenen IntSG zusammen gesetzt waren (Abbildung 3).

²<https://www.pstnet.com/eprime.cfm>

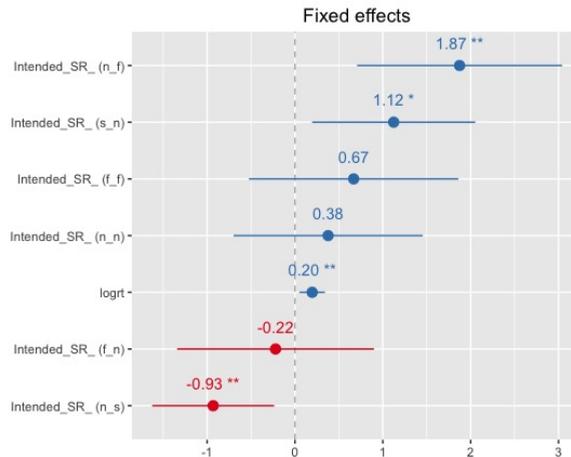


Abbildung 3 – Experiment 1: Koeffizienten (Konfidenzintervalle) des verallgemeinerten linearen gemischten Modells (GLMM; intended SR = intendierte Sprechgeschwindigkeit; n = normal; f = schnell; s = langsam).

Die Resultate dieses Experiments zeigen, dass Hörer in der Lage sind, die intendierte Sprechgeschwindigkeit zu erkennen. Wenn die intendierte Sprechgeschwindigkeit sich nicht unterscheidet, entscheiden Probanden zufällig, selbst wenn es objektive Unterschiede gibt, die über die gerade noch hörbare Unterscheidungsgrenze [12] hinausgehen. Diese Erkenntnis ist für Experiment 2 von zentraler Bedeutung.

2.2 Experiment 2

Experiment 2 dient zur genaueren Untersuchung des möglichen Einflusses von Surprisal auf die Perception von Sprechgeschwindigkeit. Um diesen Einfluss besser zu untersuchen, wurde die intendierte Sprechgeschwindigkeit konstant gehalten (nur Intervalle mit normaler Geschwindigkeit wurden untersucht). Außerdem wurden nur Intervalle von deutschen Muttersprachlern untersucht, da nur für diese die Berechnung des auf statistischen Sprachmodellen (N-grammen) basierenden Surprisal-Wertes sinnvoll vorgenommen werden kann.

2.2.1 Materialien

Die Skalierungsstudie beinhaltete insgesamt 256 Stimuli, von denen 50 Intervalle von sechs deutschen Sprechern (3 weiblich, 3 männlich) in normaler Sprechgeschwindigkeit ausgewertet wurden. Für diese Intervalle wurden die Surprisal-Werte auf der Basis eines Trigramm-Modells für das Deutsche (aus [8]) berechnet. Diese Werte wurden durch die Dauer der jeweiligen Intervalle dividiert, um einen vergleichbaren Wert für alle Intervalle zu erhalten. Die Verteilung der Surprisal/Dauer-Werte lag zwischen 0,0075 und 0,0147 (Mittelwert 0,0111 und Standardabweichung 0,0019).

2.2.2 Versuchspersonen

Insgesamt nahmen 11 Probanden (9 weiblich, 2 männlich) an Experiment 2 teil. Alle studierten an der Universität des Saarlandes in Saarbrücken. Ihr Durchschnittsalter lag bei 26,5 Jahren, keine der Personen war bereits Teilnehmer bei Experiment 1. Die Probanden erhielten für die Teilnahme am Experiment eine monetäre Entschädigung.

2.2.3 Methode

Das Skalierungsexperiment wurde in PRAAT [13] mithilfe des Demo-Fensters implementiert. Die Probanden hörten ein Intervall und mussten sich danach auf einem Balken mit den Endpunkten *langsam* und *schnell* durch Anklicken entscheiden, wo sie den gehörten Stimulus einordnen würden. Es gab keine Zeitbegrenzung.

2.2.4 Ergebnisse

Die vertikale Einordnung auf dem Balken wurde als (kontinuierliches) Maß für die wahrgenommene Sprechgeschwindigkeit verwendet. Ein Lineares Gemischtes Modell mit wahrgenommener Geschwindigkeit als abhängiger Variable, Sprecher, Surprisalrate, LabSG und deren Interaktionen, sowie den Probanden als zufälliger Effekt wurde mit der JMP Software [14] berechnet. Lediglich Sprecher wurde als signifikanter Faktor identifiziert ($F(5,521)=5,476$, $p < 0,0001$). Alle anderen Faktoren und Interaktionen erreichten das Signifikanzniveau nicht.

3 Diskussion und Ausblick

Die Ergebnisse der Diskriminationsstudie zeigen, dass die Hörer (L1: Deutsch) in der Lage waren, die intendierten Sprechraten sowohl der deutschen als auch der französischen Stimuli voneinander zu unterscheiden. Die Diskrimination gleich schnell intendierter Stimuli (z.B. normales Tempo vs. normales Tempo) war für die Versuchspersonen ungleich schwerer (zwischen 16% und mehr als 20% weniger korrekt und auf Zufallsniveau). Die tatsächlich gemessene Sprechgeschwindigkeit war für die Entscheidung der Probanden nicht entscheidend.

Die Ergebnisse von Experiment 2, in dem für die Sprechgeschwindigkeit kontrolliert und Intervalle von mehreren Sprechern verwendet wurden, bestätigte den Befund, dass tatsächliche Sprechgeschwindigkeit innerhalb einer Kategorie von intendierter Sprechgeschwindigkeit keinen entscheidenden Einfluss nimmt. Dies gilt auch für Surprisal. Hörer scheinen Sprechgeschwindigkeit nicht anders schnell wahrzunehmen, wenn sich die Surprisalwerte in einem Sprachintervall unterscheiden.

Ein Einfluss von Surprisal auf die Perception von Sprechgeschwindigkeit konnte zwar in den beiden Experimenten nicht gezeigt werden, weitreichende Aussagen scheinen an dieser Stelle jedoch verfrüht. Eine mögliche Erklärung für das Fehlen des Einflusses von Surprisal in der Perception der Sprechgeschwindigkeit liegt in der Effektstärke von Surprisal, da die Effekte von Surprisal im Vergleich zu anderen phonetischen Faktoren eher subtil sind und diese eher modulieren als überlagern. Eine größer angelegte Studie, mit weiteren Stimuli verschiedener Sprecher, die einen größeren Bereich der Variation des Surprisal abdecken, könnte hier eine gute Option darstellen.

Eine weitere Erklärung liegt in der Methode des Skalierungsexperiments in Kombination mit Sprechgeschwindigkeit. Anders als bei Bewertungsexperimenten wie in [10], bei denen ein gehörter Stimulus (im Vergleich zu anderen Stimuli) als mehr oder weniger gut passend für eine Situation bewertet wird, ging es im Skalierungsexperiment um Sprechgeschwindigkeit, bei der für jeden Stimulus individuell die gehörte Geschwindigkeit festgelegt werden musste. Eine weitere Einschränkung hinsichtlich der Generalisierbarkeit der Ergebnisse ist darin zu sehen, dass die Vorhersagbarkeit im Kontext zwar für die Sprecher bei der Produktion der Stimuli relevant gewesen sein mochte, für die Probanden, die ja nur relativ kurze Intervalle beurteilen mussten, allerdings nur bedingt zugänglich war. Die Wahrnehmung der Sprechgeschwindigkeit war daher dominant signalbasiert und gab Top-Down-Effekte wie Surprisal keine Entfaltung in gleichem Maß wie in natürlicheren Dialogsituationen.

Literatur

- [1] RIETVELD, T. und C. GUSSENHOVEN: *Perceived speech rate and intonation*. *Journal of Phonetics*, 15, S. 273–285, 1987.
- [2] KOREMAN, J.: *The perception of articulation rate*. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS XV)*. Barcelona, Spain, 2003.
- [3] KOREMAN, J.: *Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech*. *Journal of the Acoustical Society of America*, 119(1), S. 582–596, 2006.
- [4] DELLWO, V., E. FERRAGNE, und F. PELLEGRINO: *The perception of intended speech rate in english, french, and german by french speakers*. In *Proceedings of the 3rd International Conference of Speech Prosody*, S. 101–104. Dresden, Germany, 2006.
- [5] AYLETT, M. und A. E. TURK: *The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech*. *Language and Speech*, 47(1), S. 31–56, 2004. doi:10.1177/00238309040470010201.
- [6] VAN SON, R. J. J. H. und J. P. H. VAN SANTEN: *Duration and spectral balance of intervocalic consonants: A case for efficient communication*. *Speech Communication*, 47(1), S. 100–123, 2005.
- [7] COHEN PRIVA, U.: *Informativity affects consonant duration and deletion rates*. *Laboratory Phonology*, 6(2), S. 243–278, 2015. doi:10.1515/lp-2015-0008.
- [8] SCHULZ, E., Y. M. OH, Z. MALISZ, B. ANDREEVA, und B. MÖBIUS: *Impact of prosodic structure and information density on vowel space size*. In *Proceedings of Speech Prosody 2016*, S. 350–354. Boston (MA), 2016. doi:10.21437/SpeechProsody.2016-72.
- [9] JAEGER, F. T. und E. BUZ: *Signal reduction and linguistic encoding*. In E. M. FERNÁNDEZ und H. S. CAIRNS (Hrsg.), *Handbook of Psycholinguistics*. Wiley-Blackwell, Oxford, United Kingdom, accepted.
- [10] DETHLEFS, N., H. HASTIE, H. CUAYAHUITL, Y. YU, V. RIESER, und O. LEMON: *Information density and overlap in spoken dialogue*. *Computer Speech and Language*, 37, S. 82–97, 2016.
- [11] DELLWO, V., I. STEINER, B. ASCHENBERNER, J. DANCOCOVÁ, und P. WAGNER: *BonnTempo-Corpus and BonnTempo-Tools: A database for the study of speech rhythm and rate*. In *Proceedings of the 8th ICSLP*. Jeju Island, Korea, 2004.
- [12] QUENÉ, H.: *On the just noticeable difference for tempo in speech*. *Journal of Phonetics*, 35(3), S. 353–362, 2007.
- [13] BOERSMA, P. und D. WEENINK: *PRAAT: Doing phonetics by computer*. Version 5.4 edn., 2014. URL <http://www.praat.org/>.
- [14] SAS INSTITUTE INC.: *JMP, Vers. 13*. Cary, (NC), 2016.