

Eine Datenbank für deutsche Sprache mit Lombard-Effekt

Stefanie Köster, Christoph Pörschmann, Jürgen Walter

Institut für Kommunikationsakustik, Ruhr-Universität Bochum
<http://www.ika.ruhr-uni-bochum.de>

Einführung

In Zukunft werden Synthesysteme zunehmend in einer lärmbehafteten Umgebung genutzt, z.B. bei Auskunftssystemen auf öffentlichen Plätzen, Messen o.ä. In solchen Situationen sollte es möglich sein die Sprechweise der künstlichen Stimme an die Umgebung anzupassen, um ihre Qualität zu erhöhen. Dabei kann die Synthese die menschliche Sprechweise, den sog. Lombard-Effekt [LOMBARD 1911], imitieren. Zur Analyse dieses Effekts und den daraus resultierenden Konsequenzen für die Ansteuerung der Synthese soll deshalb eine Datenbank für Lombard-Sprache aufgenommen werden. Neben der Auswahl der Texte, der Sprecher u.a. mußte hierbei besonderer Wert auf die Aufnahmeart gelegt werden. Der Sprecher muß sich in einer lärmgefüllten Umgebung befinden, die Hintergrundgeräusche dürfen allerdings nicht mit der Sprache aufgezeichnet werden. Für in der Literatur bekannte Aufnahmen von Lombard-Sprache wurde deshalb ein Kopfhörer zur Einspielung der Hintergrundgeräusche benutzt. Jedoch ist bekannt, daß sich die Wahrnehmung des Sprechers durch das Tragen des Kopfhörers ändert [PÖRSCHMANN 1998]. Es stellt sich nun die Frage, welchen Einfluß die veränderte Wahrnehmung auf den Lombard-Effekt hat, und wie man diesen Einfluß ggf. vermindern kann.

Zur Untersuchung der Wahrnehmungsstörung wurde Lombard-Sprache zunächst auf herkömmliche Art mit Kopfhörer aufgezeichnet. Eine zweite Aufnahme wurde mit einer Kompensation der Einführungs-dämpfung des Kopfhörers durchgeführt [PÖRSCHMANN 1998]. Die Sprachproben wurden sowohl anhand von Signalparametern als auch perzeptiv miteinander verglichen. Die Ergebnisse werden im folgenden präsentiert.

Datenmaterial

Für den Vergleich der Aufnahmearten wurden von sechs Sprechern, jeweils die gleichen zehn Sätze einmal ohne und einmal mit Kompensation des Kopfhörers aufgenommen. Bei den Sätzen handelt es sich um eine zusammenhängende Pressemitteilung. Über den Kopfhörer hörten die Sprecher sprachmoduliertes Rauschen mit einer Lautstärke von 80 dB.

Signalnahe Analyse

Die gewonnenen Sprachdaten wurden zunächst lautweise segmentiert. Es wurden verschiedene Parameter, die für die Beschreibung einer Sprechweise wichtig sind, untersucht.

Die Lautdauer änderte sich im Mittel über alle Laute nur geringfügig. Bei näherer Betrachtung zeigen sich für die verschiedenen Lautarten allerdings große Unterschiede. Während die Dauer von Vokalen um ca. 7 v.H. und die Dauer von Diphtongen sogar um 40 v.H. für die Aufnahmen mit Kompensation sinkt, erhöht sich die Dauer der Nasale um etwa 5 v.H. gegenüber der Aufnahme ohne Kompensation. Für die Grundfrequenz ist insgesamt eine leichte Anhebung von 5 v.H. festzustellen, hier sind die größten Veränderungen wieder bei den Diphtongen zu beobachten. Die Energie der Laute wurde in den Frequenzgruppen nach Zwicker und Feldtkeller berechnet. In allen Frequenzbändern liegt die Energie für die Aufnahmen mit Kompensation niedriger. Gleichzeitig verschiebt sich mit der Kompensation der spektrale Schwerpunkt zu höheren Frequenzen. Dies ist auch an den einzelnen Frequenzgruppen zu erkennen. Bei dem ersten Formanten ist eine um ca. 4 v.H. niedrigere Frequenz für die Aufnahmen mit Kompensation zu beobachten.

Die Ergebnisse der lautweisen Signalanalyse ähneln von der Tendenz her stark den Ergebnissen, die Jean-Claude Junqua als Merkmale des Lombard-

Effekts gegenüber ‚neutraler‘ Sprache ermittelt hat: eine Anhebung der Grundfrequenz, Anstieg der Vokaldauer und eine Verschiebung der ersten zwei Formanten [JUNQUA 1996]. Allerdings widerspricht die Verschiebung des spektralen Schwerpunkts zu niedrigeren Frequenzen seiner Beobachtung. Junqua stellte eine Erhöhung der Energie in höheren Frequenzbändern fest.

Perzeptive Überprüfung

Nachdem im Sprachsignal Unterschiede zwischen den beiden Aufnahmearten festgestellt wurden interessiert nun, ob sich diese auch perzeptiv bemerkbar machen. Es wurde ein Hörversuch durchgeführt, in dem nach der Authentizität der Stimuli gefragt wurde. Die Versuchsperson wurde zunächst mit der Situation vertraut gemacht: In einer lärmerfüllten Umgebung, z.B. in einem Bahnhof, werden ihr von einem Vorleseautomaten Nachrichten vorgelesen. Das laute Hintergrundgeräusch wird durch ein sprachsimulierendes Rauschen repräsentiert. Der Vorleseautomat wurde durch einen Lautsprecher dargestellt. Im Test wurden der Versuchsperson 28 Satzpaare aus dem aufgenommenen Datenmaterial vorgespielt. Die Versuchsperson entschied, welcher Stimulus eines Paares als der Situation angemessener klingt. Die Stimulipaaire bestanden jeweils aus den zwei unterschiedlichen Realisierungen eines Satzes desselben Sprechers. Der Pegel der Hintergrundgeräusche war 80 dB.

Es nahmen 17 Personen an dem Versuch teil. Die Auswertung des Tests ergab einen signifikanten Unterschied zwischen den beiden Aufnahmevarianten. Das Signifikanzniveau als kumulierte Wahrscheinlichkeit der binomialverteilten Zufallsvariablen „Urteil der Versuchsperson“ (0 - Vorzug ohne Kompensierung, 1 - Vorzug mit Kompensierung) lag bei 0.95, bzw. die Wahrscheinlichkeit, daß im Mittel die Kopfhörerkompensation bevorzugt wird, betrug 95 v.H. Allerdings lag der Mittelwert der vorliegenden Stichprobe nur bei 0.541, also nur knapp über der Erwartung bei Raten.

Weitere Analysen der Testergebnisse zeigten keine Abhängigkeiten der Bewertung vom Sprecher, der Versuchsperson oder von der Struktur der Sätze.

Diskussion und Ausblick

Die in der Signalanalyse festgestellten Unterschiede zwischen den Sprachaufnahmen mit Kompensation der Einführungsdämpfung des Kopfhörers und ohne

Kompensation weisen einige Gemeinsamkeiten mit den Untersuchungen von Junqua zum Unterschied zwischen Lombard-Sprache und Sprache aufgenommen im neutralen Raum auf. Dies legt den Schluß nahe, daß der Lombard-Effekt durch das Tragen des Kopfhörers verstärkt wird, ähnlich wie es graduelle Unterschiede des Effektes bei verschiedenen Lautstärken oder Arten des Hintergrundgeräusches gibt. Dafür spricht auch, daß der Mittelwert der Urteile im Hörversuch nur knapp über 0.5 liegt.

Allerdings berichtet Junqua über eine Verschiebung des spektralen Schwerpunkts zu höheren Frequenzen hin, während in dieser Arbeit eher eine Verschiebung des Schwerpunkts zu niedrigeren Frequenzen für die Aufnahmen ohne Kompensation festgestellt werden konnte. Auch die Auswertung des Hörtests zeigt, daß bei genügend großer Stichprobe die Aufnahmen mit Kompensierung im Mittel bevorzugt werden. Dies ist ein Hinweis dafür, daß das Tragen des Kopfhörers zwar ähnliche, aber dennoch nicht die gleichen Effekte wie Hintergrundgeräusche beim Sprechen provoziert.

Es ist allerdings zu vermuten, daß der Einfluß des Kopfhörers auf den Lombard-Effekt eine im Vergleich zu den Einflüssen der Art und Lautstärke des Hintergrundgeräusches sowie der Sprechervariabilität in den Hintergrund tritt, was in einem nächsten Schritt zu überprüfen bleibt.

Anmerkung

Die vorliegenden Untersuchungen wurden im Rahmen des Verbundvorhabens VERBMOBIL vom Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF) gefördert.

Literatur

- [JUNQUA 1996] JUNQUA, JEAN-CLAUDE (1996). *The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex*. 20:13–22.
- [LOMBARD 1911] LOMBARD, E. (1911). *Le signe de l'élévation de la voix*. Ann. Maladies Oreille, Larynx., Nez., Pharynx., (37):101–119.
- [PÖRSCHMANN 1998] PÖRSCHMANN, C. (1998). *Eigenwahrnehmung der Stimme in virtuellen auditiven Umgebungen*. In: *Fortschr. Akust. DAGA*, Oldenburg. Dtsch. Ges. Akust.