

Akustisch-phonetische Aspekte von Lombard-Sprache für verschiedene Sprechstile

Stefanie Köster

Institut für Kommunikationsakustik
Ruhr-Universität Bochum
koester@ika.ruhr-uni-bochum.de

Zusammenfassung

In dieser Arbeit werden signalnahe Unterschiede zwischen Lombard-Sprache und neutraler Sprache beschrieben. Dabei werden Veränderungen einiger wichtiger prosodischer und spektraler Parameter für drei verschiedene Sprechstile bzw. Textarten verglichen: einzelne Wörter, phonologisch ausbalancierte Sätze und Spontansprache.

1. Einleitung

Ein Vorteil von Systemen mit sprachgesteuerter Mensch-Maschine-Schnittstelle ist die Tatsache, dass der Benutzer Information einfach und schnell zuführen kann und zugeführt bekommt. Dies ist besonders für sog. ‚hands-busy-eyes-busy‘ Situationen der Fall, z.B. beim Navigationssystem im Auto. Gerade in solchen lauten Umgebungen tritt beim Benutzer der bekannte Lombard-Effekt auf, ein unbewusster Wechsel der Sprechweise. Für Spracherkennung ist es wichtig, sich an die veränderte Sprechweise adaptieren zu können, da die Erkennungsleistung sonst stark sinkt. Für die Sprachsynthese wiederum kann es eine geeignete Anpassung an die Umwelt sein, die Sprechweise Lombard-Sprache zu simulieren, um evtl. eine bessere Verständlichkeit zu erlangen. Eine Anpassung setzt eine genaue Analyse der Sprechweise voraus.

Im ersten Teil dieser Arbeit wird die für die Analyse benötigte Datenbank beschrieben. Anschließend wird über die Ergebnisse einer statistischen Analyse der Sprachdaten bzgl. relevanter Sprachsignal-Parameter berichtet. Für die Adaptation der Sprachsynthese ist die Wirkung der Lombard-Sprache in Bezug auf die Verständlichkeit in lauter Umgebung wichtig. Hierzu wurde ein Hörtest durchgeführt, dessen Ergebnisse im letzten Teil dargelegt werden.

2. Die Sprachdatenbank

Die erstellte Datenbank besteht aus deutscher, neutraler und mit Lombard-Effekt gesprochener Sprache. Es wurden für jeden der 3 männlichen und 3 weiblichen Sprecher drei verschiedene Sprechstile aufgenommen. Die Einzelwörter bestehen aus Anweisungen für Roboter, unter denen sich auch Zahlwörter befinden. Pro Sprecher und Sprechweise wurden 26 Einzelwörter aufgenommen. Die Spontansprache ist mit Hilfe einer Dialogvorgabe [1] erzeugt worden, die einen Dialog zur Zugreservierung

simuliert. Darüber hinaus wurden pro Sprecher und Sprechweise 10 phonologisch ausbalancierte Sätze aufgenommen.

Die Aufnahmen wurden im institutseigenen reflexionsarmen Raum durchgeführt. Für die Erzeugung der Lombard-Sprache wurde den Sprechern diffuses, binaural aufgenommenes Rauschen über Kopfhörer dargeboten [2].

3. Akustisch-phonetische Analyse

Für viele Sprachsignal-Parameter konnten signifikante Unterschiede zwischen den Sprechweisen neutrale Sprache und Lombard-Sprache festgestellt werden.

3.1 Segmentale Dauer

Für die Berechnung der Dauer wurden Pausen im Sprachsignal nicht berücksichtigt. Die mittlere Wortdauer steigt für den Dialog kaum (8%), während der stärkste Anstieg der mittleren Wortdauer bei den Einzelwörtern zu verzeichnen ist (15.8%; Sätze 14.8%). Auch auf Lautebene sind die Veränderungen der Dauer für die Spontansprache nicht signifikant (+0.5%), während die Lautauern für die Sätze (12.6%) und die Wörter (16.6%) signifikant ansteigen.

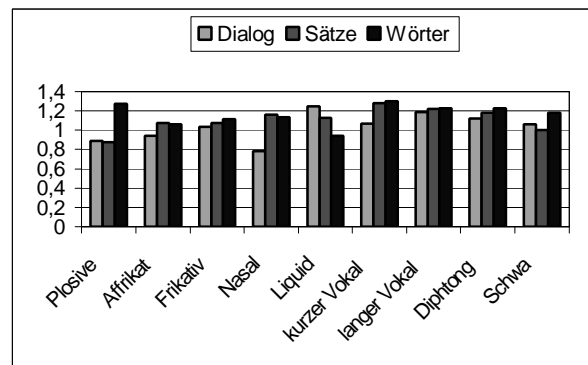


Bild 1: Verhältnis der Wortdauer Lombard/ neutrale Sprache

Bild 1 zeigt die Verhältnisse der Lautauern von Lombard-Sprache zu neutraler Sprache für die einzelnen Lautklassen. Auffällig ist die starke Zunahme der Plosiv-Lautauern für die Einzelwörter, während diese für die Dialoge und Sätze stark abnimmt. Die Verschlusspause geht hier nicht mit in die Berechnung ein. Für Vokale ist die Position der Silbe für den Grad der Länge entscheidend. Sowohl für die Dialoge als auch für die Sätze erhöht sich die Dauer der Vokale in Silben vor

Pausen um über 60%, für Silben am Satzende beträgt die Steigerung nur 34% (Dialog) bzw. 28%. Dies deutet auf eine starke Akzentuierung der Phrasenenden hin.

3.2 Grundfrequenz

Für Lombard-Sprache liegt die mittlere Grundfrequenz signifikant höher als für neutrale Sprache. Dies ist für alle drei Textarten zu beobachten (s. Bild 2). Auffällig ist hier der große Unterschied in der Varianz der Grundfrequenz für die Sätze, der auf eine größere Variation der Grundfrequenz hindeutet. Dieser Trend ist allerdings auch in geringerem Maße für den Dialog und die Wörter zu beobachten.

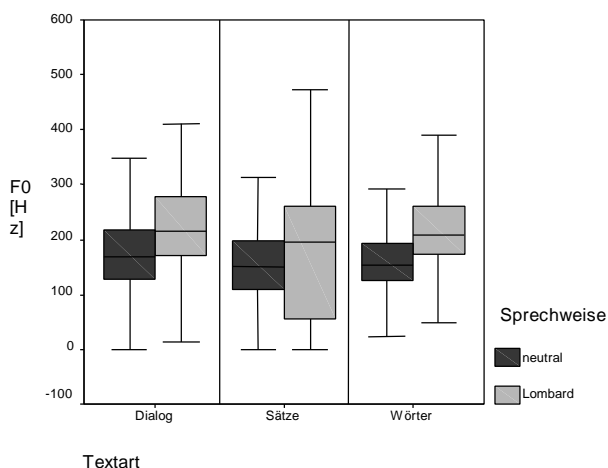


Bild 2: Mittlere Grundfrequenz für die verschiedenen Textarten

3.3 Formanten und Formant-Bandbreiten

Während die mittlere Frequenz des ersten Formanten für alle drei Textarten signifikant steigt, verändern sich die 2. und 3. Formanten kaum. Die Entwicklung der Bandbreiten ergibt ein uneinheitliches Bild.

	F1		F2		F3	
	F1	BBreite	F2	BBreite	F3	BBreite
Dialog	18,61	-3,49	3,28	-21,99	1,98	-24,17
Sätze	26,5	21,65	0,29	31,01	-2,21	28,46
Wörter	15,3	5,97	1,53	-8,92	0,76	-10,66

Tabelle 1: Mittlere Formantfrequenz- und Formantbandbreitenänderung in Prozent

Während für Dialog und Wörter die Formant-Bandbreiten eher schmaler werden, verbreitern sich die Frequenzbänder für die Sätze. Die Verschiebung des 1. Formanten ist auch in Bild 3 zu beobachten, das die Formantfrequenzen der langen Vokale in der F1-F2 Ebene zeigt.

3.4 Spectral Tilt und spektraler Schwerpunkt

Für alle drei Textarten lassen sich sowohl beim spectral tilt als auch beim spektralen Schwerpunkt die gleichen Tendenzen beobachten. Die Flanke des LPC-Spektrums flacht für Plosive, Affrikate und Frikative ab, während sie für alle anderen Lautklassen steiler wird. Das Abflachen des spectral tilt tritt sehr ausgeprägt bei den Sätzen auf, während die größere Steigung des Spektrums viel stärker bei Wörtern und dem Dialog zu beobachten ist.

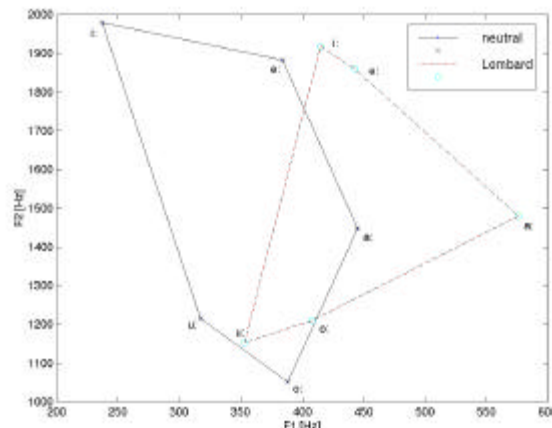


Bild 3: Formantfrequenzen in der F1-F2 Ebene für lange Vokale

Der spektrale Schwerpunkt verschiebt sich für alle Lautklassen zu höheren Frequenzen hin, die größte Verschiebung tritt hier bei Vokalen und Diphtongen auf, während fast alle Konsonanten, ausgenommen die Nasale, eine geringere Verschiebung des spektralen Schwerpunkts aufweisen.

4. Hörversuch Worterkennung

Um die Frage zu beantworten, ob der Lombard-Effekt tatsächlich zu einer höheren Verständlichkeit führt, wurde folgender Hörversuch durchgeführt. Aus der Datenbank wurden einsilbige Einzelwörter ausgesucht und im Sprachpegel normiert. Den Versuchspersonen wurden die Einzelwörter in dem mit diffusem Rauschen (SNR 5dB) beschallten reflexionsarmen Raum des Instituts vorgespielt mit der Anweisung, das Gehörte aufzuschreiben. Insgesamt wurden pro Sprechweise 89 Wörter dargeboten, 20 Versuchspersonen nahmen an dem Versuch teil. Es wurden die Worterkennungsraten für neutrale und Lombardsprache ausgewertet. Die Lombard-Sprache erreichte trotz des normalisierten Pegels eine Worterkennungsraten von 92% gegenüber 75% für neutrale Sprache.

5. Zusammenfassung und Ausblick

Es lassen sich einige signifikante Unterschiede zwischen Lombard-Sprache und neutraler Sprache sowohl für prosodische als auch für spektrale Parameter beobachten. Die Unterschiede der Sprechweisen führen zu einer höheren Verständlichkeit der Lombard-Sprache in Lärm. Interessant ist nun ein Vergleich von Lombard-Sprache zu bewusst deutlicher Sprache. Entsprechende Untersuchungen werden bereits vorgenommen.

6- Literatur

- [1] ITU-T Contribution COM 12-35. Development of scenarios for a short conversation test. Source: Federal Republic of Germany (S. Möller). International Telecommunication Union, CH-Geneva, 1997
- [2] Köster, S., Pörschmann, C., Walter, J. Eine Datenbank für deutsche Sprache mit Lombard-Effekt. Fortschritte der Akustik - DAGA 2000, Oldenburg, 2000