

Simulation des emotionalen Ausdrucks ‚Freude‘ mit Sprachsyntheseverfahren

Felix Burkhardt, Walter F. Sendlmeier

*TU-Berlin, Institut für Kommunikationswissenschaft, Einsteinufer 17, 10587 Berlin
felixbur | sendl@kgw.tu-berlin.de*

Zusammenfassung: Es wurde der Einfluß akustischer Parameter auf den emotionalen Ausdruck von Sprachäußerungen, speziell Freude, untersucht. In einer Voruntersuchung wurden neun Satzpaare analysiert. Ausgehend von den Ergebnissen der Analyse wurde ein neutraler Satz hinsichtlich der Parameter Lage der Formanten, Form der Grundfrequenzkontur und der Dauern der Laute systematisch durch PSOLA- und LPC-Verfahren variiert. Ein Hörversuch ergab signifikante Effekte aller drei Parameter sowie der Interaktion zwischen Grundfrequenzkontur und Lautdauern.

EINLEITUNG

Hörer sind grundsätzlich in der Lage, aufgrund lautsprachlicher Indikatoren den emotionalen Gehalt einer Äußerung und den emotionalen Zustand eines Sprechers zu beurteilen. Es ist jedoch bislang weitgehend ungeklärt, welche akustischen Korrelate primär für die emotionale Färbung verantwortlich sind.

Die Simulation emotionaler Sprechweise ist seit den Fortschritten auf dem Gebiet der Text-to-Speech Synthese von besonderem Interesse, da durch sie die Natürlichkeit synthetisierter Sprache und damit auch ihre Einsatzmöglichkeiten stark verbessert werden könnten.

Die Emotion ‚Freude‘ wurde im Gegensatz zu anderen Basisemotionen wie ‚Trauer‘, ‚Ärger‘ oder ‚Angst‘ in früheren Perzeptionsexperimenten von Hörern eher schlecht erkannt (Bergmann et al. 1988). Aus diesem Grund wurden ihre akustischen Korrelate in dieser Untersuchung gesondert betrachtet.

In einer Voruntersuchung wurden 9 Satzpaare analysiert, die aus einer im Rahmen eines DFG-Projektes erstellten Sprachdatenbank stammen. Die neutralen und freudigen Äußerungen wurden von Schauspielern simuliert. Die akustische Analyse erbrachte systematische Unterschiede zwischen den neutralen und den freudigen Äußerungen. Einige dieser akustischen Merkmale wurden in der vorliegenden Arbeit mithilfe systematisch variiert Stimuli in einem Hörversuch auf perzeptive Relevanz überprüft.

GENERIERUNG DER STIMULI

Ausgangsmaterial für die Resynthese waren eine neutral und eine freudig realisierte Äußerung aus der oben erwähnten Datenbank. Untersuchungsgegenstand waren Formantlagen, Lautauern und Intonationskonturen.

Die Formanten wurden auf zwei Arten mittels LPC-Resynthese verändert.

- In Version 1 wurde der erste Formant um 6 % angehoben und der zweite um 4 %. Dies entspricht in etwa den Mittelwerten, die sich aus der Voruntersuchung ergeben hatten.
- In der zweiten Version wurden die ersten fünf Formanten um 5 % erhöht. Diese Manipulation wurde hinzugenommen, weil die erste sich im akustischen Eindruck nicht sehr stark bemerkbar machte. Zudem liegt die physiologische Begründung für das Anheben der Formanten in einer Verkürzung des Ansatzrohres (Laver 1991), und die betrifft das gesamte Spektrum.

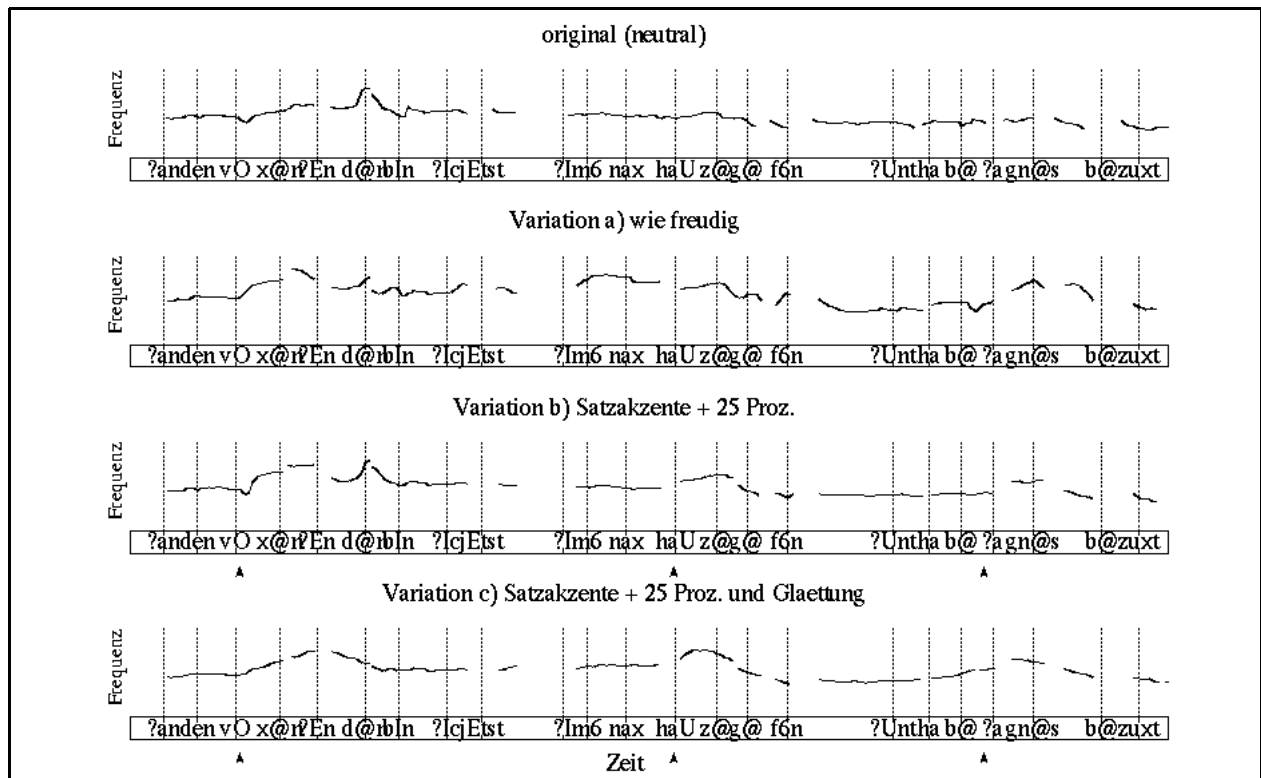


Abbildung 1: Die verwendeten Grundfrequenzkonturen

Der Grundfrequenzverlauf wurde auf drei Arten durch TD-PSOLA-Resynthese manipuliert.

- Die Intonationskontur wurde der freudigen Version nachgebildet. Da die TD-PSOLA-Resynthese Artefakte bei starken F0-Manipulationen erzeugt, wurde der gesamte Verlauf dann auf die ursprüngliche Stimmlage abgesenkt.

- In dem zugrundeliegenden Satz wurden drei Phrasenbetonungen angenommen; die drei betonten Silben wurden hinsichtlich der Grundfrequenz um 25 % angehoben. Eine stärkere Anhebung der betonten Silben bei freudiger Sprechweise hatte sich aus der Voruntersuchung ergeben.
- Als dritte Variation wurden die drei Satzbetonungen um 50 % angehoben. Um unnatürliche Sprünge zu vermeiden, wurde 2-5 Silben vor und nach diesen Zielpunkten hinsichtlich der Grundfrequenz linear angehoben bzw. abgesenkt.

Die originale Intonationskontur sowie die drei manipulierten sind in Abbildung 1 dargestellt.

Die Dauern wurden auf zwei Arten verändert:

- Im einen Fall wurden die Lautdauern durch TD-PSOLA-Resynthese denen der freudigen Version angepasst.
- In der zweiten Variante wurden die drei phrasenbetonenden Silben um 24 % verlängert und alle anderen um 20 % verkürzt.

DURCHFÜHRUNG DES HÖRTESTS

Insgesamt gab es einschließlich der neutralen Versionen drei Varianten für die Formantstruktur, vier für die F0-Kontur und drei für die Dauern. Hinzu kam noch eine resynthetisierte Version des freudigen Satzes, dessen Grundfrequenz und Dauern der neutralen Version nachgebildet waren. Dies ergab eine Stimulusmenge von $3 * 4 * 3 + 1 = 37$ Sätzen. Diese wurden in zufälliger Reihenfolge je zwei Mal auf ein DAT-Band gespielt.

Die Versuchspersonen bekamen einen Fragebogen und wurden mündlich instruiert, die Frage "Wie freudig klingt die Sprecherin?" mit einem Kreuz auf einer siebenstufigen Rating-Skala zwischen "gar nicht" und "sehr" zu beantworten. Teilnehmer waren 10 Frauen und 7 Männer im Alter von 23-35 Jahren. Die Stimuli wurden mittels eines tragbaren DAT-Rekorders und Kopfhörern dargeboten.

ERGEBNISSE

Die Ergebnisse wurden einer Varianzanalyse mit kompletter Messwiederholung unterzogen. Alle drei Haupteffekte waren hoch signifikant (p-Wert unter 0.01). Die Interaktion zwischen F0-Kontur und Dauern wurde ebenfalls signifikant (p-Wert von 0.015). Alle anderen Interaktionen waren nicht signifikant.

Beide Manipulationen der Formantlagen wurden als freudiger beurteilt; ein signifikanter Unterschied zwischen den Manipulationen ergab sich jedoch nicht. Die Version, bei der die ersten fünf Formanten um 5 % angehoben waren, wurde allerdings tendenziell besser bewertet. Dies liegt vermutlich daran, dass freudige Sprechweise sich generell durch mehr Energie in den höheren Frequenzbereichen auszeichnet (Klasmeyer und Sendlmeier 1999).

Die F0-Kontur mit den um 25 % angehobenen Satzbetonungen (Variation b)) hat die höchste Bewertung als freudige Sprechweise erfahren, gefolgt von der Variation c). Die

F0-Kontur, die vollständig der freudigen Version nachgebildet wurde, wurde nicht als freudiger beurteilt als der neutrale Verlauf.

Das der freudigen Version nachempfundene Dauermodell wurde ebenfalls nicht als freudiger beurteilt. Lediglich Variation b) (betonte Silben um 24 % gedehnt, andere um 20 % gekürzt) erbrachte eine höhere Bewertung.

Die Bewertungen der Intonationskonturen in Abhängigkeit von den Dauermodellen ergab, dass es für die neutrale Grundfrequenzkontur und Intonationskontur c) keine Rolle spielt, welches Dauermodell verwendet wird. Die Intonationskonturen b) und d) wurden unter der zweiten Variation der Lautauern besser bewertet.

Die freudige Version, deren Dauern und Intonationskontur denen der neutralen angepasst waren, wurde von allen Stimuli als am freudigsten klingend bewertet.

DISKUSSION

Einige der durchgeführten Manipulationen der ursprünglich neutralen Version haben tatsächlich zu einer Attribution der Stimuli als freudiger geführt.

Die Beurteilung des freudigen Satzes mit neutraler Prosodie als freudig und die Beurteilung des neutralen Satzes mit freudiger Prosodie als nicht so freudig weist auf die zusätzliche Relevanz anderer stimmlicher und artikulatorischer Merkmale hin.

Der Haupteffekt, der sich für die zweite Variation der Dauern ergab, bestärkt die These, dass freudige Sprechweise durch stärkere Betonung gekennzeichnet ist (Murray, Arnott 1993). Die Verlängerung der betonten und die Verkürzung aller anderen Silben führte vor allem dazu, dass der Satzrhythmus prägnanter wurde.

LITERATUR

- Bergmann G., Goldbeck T. & Scherer K. R. (1988): "Emotionale Eindrucks Wirkung von prosodischen Sprachmerkmalen", *Zeitschrift für experimentelle und angewandte Psychologie* **35**, S. 167-200
- Klasmeyer, G. und Sendlmeier, W. F. (1999): "Voice and Emotional States". In: R. Kent & M. Ball (eds.) *The Handbook of Voice Quality Measurement*. Singular Publishing Group, San Diego, in Druck
- Laver, J. (1991): *The Gift of Speech*, Edinburgh, Edinburgh University Press 1991, S. 221
- Murray, I. R. und Arnott, J. L. (1993): "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion", *JASA* 1993, Vol. **2**, S. 1097-1107