

PERZEPTIVE UNTERSUCHUNGEN ZUR ZEITSTRUKTUR VON /g/-LAUTEN

Jörg Helbig, Technische Universität Dresden

1. Einführung

Die Explosivlaute oder Plosive stellen besonders markante Erscheinungen im Sprachsignal dar. Sie zeichnen sich durch schnelle Änderungen sowohl in zeitlicher als auch in spektraler Hinsicht aus. Seit den fünfziger Jahren wurden zahlreiche Untersuchungen zu dieser Lautgruppe durchgeführt, aus denen ein umfangreiches Merkmalbündel zur Beschreibung von Artikulationsart (stimmhaft oder stimmlos) und Artikulationsort (labial, alveolar/dental oder velar/palatal) der Plosive hervorgegangen ist (/1/, /2/, /3/).

Häufig wird dabei die Frage diskutiert, ob für Perzeption und Klassifikation der Plosive die zeitlichen oder die spektralen Merkmale dominant sind. Die Antwort darauf tendiert zu einer größeren Wertigkeit der Spektralmerkmale für alle Plosive. Mit der im folgenden vorgestellten Untersuchung soll jedoch am Beispiel der stimmhaften Plosive gezeigt werden, daß sich diese Frage der Merkmaldominanz nicht pauschal beantworten läßt, sondern im Zusammenhang mit den Artikulationsstellen zu betrachten ist (/4/), und daß zeitliche Ereignisse eine starke Signalwirkung für die weitere Verarbeitung ausüben können.

Es ist bekannt, daß sich velare Verschußlaute von den übrigen durch eine längere Voice-onset-time und z.T. auch durch eine größere Burstamplitude abheben (/5/, /6/, /7/, /8/, /9/). In einem Hörtest sollte untersucht werden, ob diese im wesentlichen zeitlichen Merkmale, die artikulatorisch bedingt und somit ohnehin vorhanden sind, auch für die Perzeption velarer Plosive eine besondere Rolle spielen.

2. Material

Sechs natürlich gesprochene /g/-Vokal-Verbindungen und eine /g/-Konsonant-Verbindung, die bei auditiver Kontrolle als deutlich artikuliert bezeichnet werden konnten, dienten als Untersuchungsmaterial. Davon stammten drei aus den Logatomen /aga/, /igi/ und /ugu/ des Sprechers 1 und die restlichen aus den in fließend gesprochenen Sätzen vorkommenden Wörtern /gut/, /Gas/, /Giebel/ und /Glut/ eines zweiten Sprechers. Diese Muster wurden digitalisiert (Abtastfrequenz 20 kHz, Tiefpaß 8 kHz) und anhand des Zeitsignals manipuliert. Bild 1 zeigt die dabei verfolgte Vorgehensweise.

Zunächst wurden 7 Originalstimuli herausgeschnitten (Schnittpunkt T1), deren extrahierter Zeitabschnitt von der Verschußpause bis zum Ende des Folgevokals bzw. bis nach das /t/ von /Glut/ reichte. Danach wurden diese Originale um den Burstabschnitt des /g/-Lautes verkürzt. Es wurde versucht, den Schnitt kurz vor Beginn des Stimmeinsatzes zu legen (Schnittpunkt T2). Die Wichtung der Zeitfunktion mit einem speziellen Zeitfenster zur Vermeidung abrupter Einsätze war nicht erforderlich, da der Schnitt in Zeitpunkten minimaler Signalenergie erfolgte.

3. Hörergruppe und Testbedingungen

Das Testmaterial wurde von 29 Hörern beurteilt. Alle waren nicht phonetisch geschulte, deutsche Muttersprachler, stammten jedoch aus verschiedenen Sprachräumen (hauptsächlich Berlin und Sachsen). Ihnen wurden die 14 Stimuli je dreimal mit einer Sekunde Pause und einer Antwortzeit von ca. 10 Sekunden angeboten. Die Testpersonen wurden aufgefordert, die Stimuli in lateinische Buchstaben zu transkribieren und für jeden Stimulus (notfalls durch raten) genau eine Entscheidung anzugeben. Ihnen wurde lediglich mitgeteilt, daß es sich dabei um Lautverbindungen oder kurze Wörter handelte, die aus natürlicher Sprache herausgeschnitten worden waren. Es gab keinerlei Informationen über das Versuchsziel. An die Vorführbedingungen wurden ebenfalls keine besonderen Anforderungen gestellt. Der Test wurde unterschiedlich großen Hörergruppen mit Hilfe handelsüblicher Kassettengeräte in verschiedenen Räumen zur Beurteilung angeboten.

4. Auswertung

Da die Hörer nicht zwischen vorgegebenen, feststehenden Kategorien zu wählen hatten, mußte das breite Spektrum der Urteile dem Versuchsziel entsprechend geordnet werden. In diesem Sinne waren die richtige Zuordnung der jeweiligen Vokale und auch die vollständige Transkription des Wortes /Glut/ (mit oder ohne /l/, Vokal /u/ oder /o/, mit oder ohne /t/ am Ende) nebensächlich. Die Urteile wurden lediglich entsprechend ihren Anfängen in die 5 Gruppen:

1. initiales /g/
2. initiales /b/
3. initiales /d/
4. initialer Vokal
5. sonstige geordnet.

So fielen z.B. die Urteile /goud/ und /knut/ für /Glut/ in die Gruppe 1 und /bot/ in die Gruppe 2. Als stimmlose Plosive beurteilte Initiallaute wurden den entsprechenden stimmhaften Plosiven mit gleichem Artikulationsort zugeordnet. Das war jedoch nur in 11 der insgesamt 406 Urteile der Fall. Die einzigen Vertreter für die Gruppe 5 (sonstige) waren ein /hu/ für /gu/ mit Burst sowie ein /lut/ für /Glut/ ohne Burst.

5. Ergebnisse

Die in Bild 2 dargestellten Ergebnisse zeigen, daß zunächst die Originalstimuli mit Burst wie erwartet nahezu zweifelsfrei der Kategorie "initiales /g/" zugeordnet wurden. Dieser /g/-Eindruck geht jedoch bei den Stimuli ohne Burst stark zurück und verlagert sich kontextabhängig in Richtung /b/ oder /d/, ohne daß dabei die spektralen Eigenschaften des Vokaleinsatzes verändert werden.

Der Kontext geschlossener Vokale wie /u/ erzeugt einen /b/-Eindruck. Das gilt möglicherweise auch für Fälle mit Konsonanten vor dem Vokal, läßt sich am verwendeten Einzelbeispiel /Glut/ nur vermuten.

Der Kontext offener Vokale wie /a/ und /i/ ruft einen /d/-Eindruck hervor. Die relativ große Anzahl der Urteile "initialer Vokal" bei Kontext /i/ geht vollständig auf das Konto eines Stimulus, bei dem möglicherweise der Schnitt T2 schon innerhalb

des Vokalanfangs gesetzt wurde. Darüber hinaus besitzen die Verbindungen /di/ und "initiales /i/" (also Glottisschlag-/i/) sehr ähnliche Spektraleigenschaften, so daß Verwechslungen durchaus denkbar sind.

Diese Ergebnisse stützen die Vermutung einer besonderen Stellung des /g/-Lautes und seiner Burstphase gegenüber den anderen stimmhaften Plosiven /b/ und /d/. Der /g/-Burst in einem zeitlichen Mindestabstand vor dem Folgevokal wirkt gewissermaßen als /g/-Initialisierung und stellt beim Hörer die Weichen in Richtung "Eigenschaftskomplex /g/". Der Mindestabstand liegt in der Größenordnung von 20...30 ms und entspricht der u.a. in /6/, /10/ und /11/ angegebenen Zeitspanne, die notwendig ist, um zwei aufeinanderfolgende akustische Ereignisse getrennt wahrnehmen zu können. Wird dieser Abstand zu klein bzw. fehlt der Burst ganz, so wendet der Hörer "Eigenschaftskomplex /d/ oder /b/" an. Diese Wahl wiederum ist von den spektralen Eigenschaften des Folgevokals abhängig.

Obwohl die im vorgestellten Hörversuch verwendete Stichprobengröße noch keine gesicherten Aussagen erlaubt, wird der beschriebene Effekt deutlich demonstriert. Die zeitlichen und die spektralen Merkmale von stimmhaften Plosiven scheinen für unterschiedliche Artikulationsstellen verschiedene Wertigkeiten zu besitzen. Gleiche Spektralmerkmale können bei abweichender Lautumgebung unterschiedlich interpretiert werden. Die unkritischen Randbedingungen und die geringen Vorinformationen für die Hörer weisen darauf hin, daß es sich dabei um eine recht robuste Erscheinung handelt, die auch in fließender Sprache für die Perzeption relevant sein dürfte.

Literaturverzeichnis

- /1/ Halle M., Hughes G.W., Radley J.A.: Acoustic properties of stop consonants. J. Acoust. Soc. Am., vol. 29(1957) 107-116
- /2/ Blumstein S.E., Stevens K.N.: Perceptual invariance and onset spectra for stop consonants in different vowel environments. J. Acoust. Soc. Am., vol. 67(1980) 648-662
- /3/ Edwards T.J.: Multiple Analysis of intervocalic english plosives. J. Acoust. Soc. Am., vol. 69(1981) 535-547
- /4/ Seck R.: Messung von Ausgeprägtheitsfunktionen als Mittel zur Bestimmung der Wirksamkeit akustischer Merkmale bei der sprachwahrnehmung. Diss. Techn. Universität München, 1987
- /5/ Lisker L., Abramson A.S.: Some effects of context on voice-onset-time in english stops. Lang. Speech 10(1967) 1-28
- /6/ Stevens K.N., Klatt D.H.: Role of formant transitions in the voiced-voiceless distinction for stops. J. Acoust. Soc. Am., vol. 55(1974) 653-659
- /7/ Kewley-Port D.: Measurement of formant transitions in naturally produced stop-vowel syllables. J. Acoust. Soc. Am., vol. 72(1982) 379-389
- /8/ Ohde R.N.: Stop identification from vocalic transition plus vowel segments of CV and VC syllables: A follow up study J. Acoust. Soc. Am., vol. 69(1981) 297-300
- /9/ Fischer H.-U.: Untersuchungen zur Zeitstruktur von Explosivlauten. Diplomarbeit, Techn. Universität Dresden, 1990
- /10/ Blutner F.: Modell zur organischen Sprachverarbeitung.

Diss. A Techn. Universität Dresden, 1975
 /11/ Pisoni, D.B.: Identifikation und diskrimination of the relative onset time of two component tones: Implications for voicing perception in stops.
 J. Acoust. Soc. Am., vol. 61(1977) 1352-1361

ZEITFUNKTION (ATF=19.2kHz)

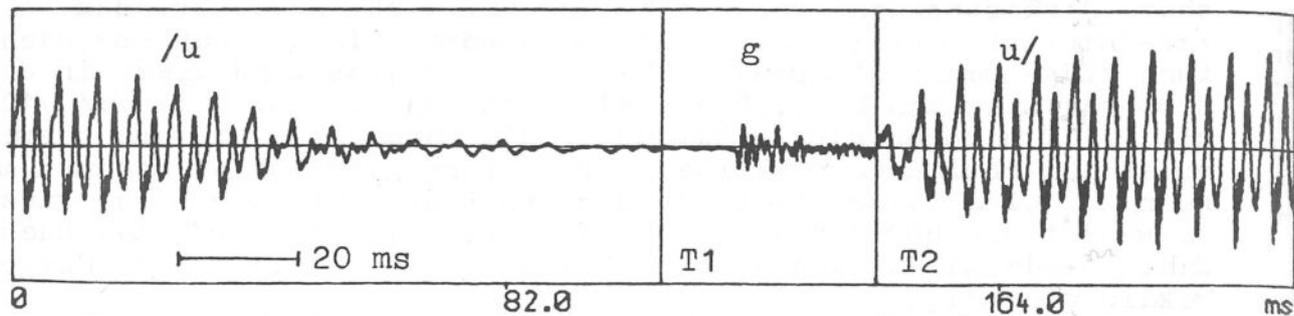


Bild 1.: Erzeugung der Stimuli am Beispiel /ugu/

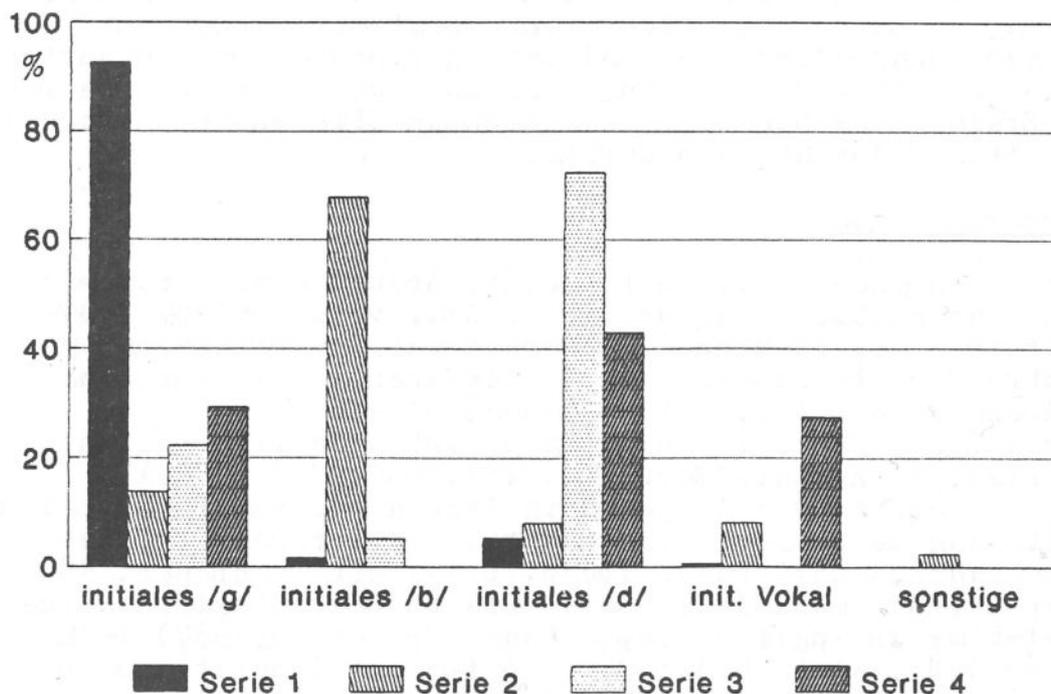


Bild 2.: Hörerurteile in Prozent

Serie 1: /g/ mit Burst, alle Kontexte
 Serie 2: /g/ ohne Burst, Kontext /u/
 Serie 3: /g/ ohne Burst, Kontext /a/
 Serie 4: /g/ ohne Burst, Kontext /i/