

PERZEPTIVE BEWERTUNG DER TSCHECHISCHEN EXPLOSIVLAUTE

Sáva Hlaváč

Phonetisches Labor
Institut für tschechischen Sprache ČSAV
Praha, ČSSR

RESUME

Die beschriebenen Experimente sollten Beziehungen zwischen signifikanten Merkmalen der tschechischen Explosivlaute feststellen. Der Wichtigkeitsgrad, gemeinsame Vertauschmöglichkeit und Reduktion der Merkmale wurden definiert. Die Merkmale waren: das explosive Geräusch, das postexplosive (PE) homorgane Geräusch und Transiente an dem die Konsonante begleitenden periodischen Signal.

EINLEITUNG

Gegenwärtige Methoden der Sprachsynthese benutzen meistens digitalisierte Segmente mit der Länge 20 + 30 ms. Im Verarbeitungsprozess muss man wichtige perzeptionsrelevante Merkmale bestimmen, die keine Verzerrung erleiden dürfen. Für Explosive ist das besonders wichtig ihres transienten Charakters wegen. Das Sprechsignal wird durch verzerrende Einflüsse beeinträchtigt; zugleich wirkt aber auch deren Kompensationsfaktor. Die Gesamtverständlichkeit kann auch dann erhalten bleiben, wenn einzelne Faktoren der Laute in veränderter Weise realisiert werden. In der Sprachsynthese kommt es darauf an, welcher Teil des akustischen Signals für die Perception des untersuchten Explosivs der wichtigste ist: das explosive Geräusch, das PE homorgane Geräusch oder Transiente an dem Konsonanten begleitenden periodischen Signal. Das Experiment wurde mit tschechischen stimmlosen Explosivlauten [p], [t], [t₁], [k] durchgeführt. Bei der Bewertung anderer Explosivlaute muß man die Verschiedenheiten in der Lautbildung und deren perzeptionellen Bewertung respektieren.

BENUTZTE MESSMETHODE

In der Sprechforschung benutzt man häufig die segmentale Synthese. Für die Analyse der Sprecherelemente ist es jedoch zweckmässig mit natürlichem Signal zu arbeiten. In unserem Labor verwendet wird die Methode der mechanischen Montage des bei höher Geschwindigkeit 76,2 cm/s arbeitenden vollspurigen Tonbandes. Mittels dieser Methode ist es möglich:

1. beliebige Kombination der Signalabschnitte zu konstruieren und einzelne Teile dabei auslassen

2. ausgewählte Abschnitte genau an Stellen zu bringen die vorher im Spektrogramm definiert wurden
3. voraus-definierte Signalübertragung zu bestimmen und das Signal in der Intensität zu modifizieren
4. die gesamte Aufnahmelänge konstant zu halten bzw. sie nach Bedarf zu modifizieren
5. einzelne Teile des Signals durch andere zu ersetzen
6. das Originalspektrum der Restteile und die Zeitverhältnisse nur minimal beeinträchtigen.

Die Zeitdauer des Überganges der verbundenen Teile war meistens 10 ms (Abb. 1).

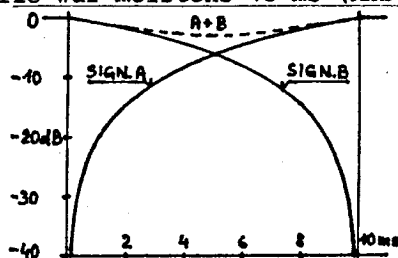


Abb. 1

Diese Länge ermöglichte die maximale Information über das ursprüngliche Signal zu behalten und gleichzeitig das Entstehen von Übergangserscheinungen zu begrenzen. Der Übergang wurde dort gelegt, wo konnte man keine signifikante Signaländerung erwarten.

Das Material wurde wie folgt verarbeitet:

- a) Das Spektrogramm der Aufnahme wurde fertig.
 - b) Nach Auswerten des Spektrogramms wurde mittels Montage ein modifiziertes Signalmuster erstellt.
 - c) Das Spektrogramm des Musters wurde fertig.
 - d) Das Muster wurde perzeptiv gewertet.
- Die Bewertung der Ergebnisse ist dadurch beeinflusst, daß das Maß an Wichtigkeit des perzipierten Signals in keinem konstanten, linearen Verhältnis zur Intensität steht. Die genaue Auswertung der Zeit- und Spektralverhältnisse muß also sehr sorgfältig ausgeführt werden.

GRAPHISCHE DARSTELLUNG UND PHYSIOLOGISCHE PROZESSEN

Unter Anwendung der Booleschen Algebra können die, für die Identifizierung des tschechischen Explosivs notwendigen Bedingungen

wie folgt definiert werden

$$E_n = \bar{F}_0 F_{\text{char}} T$$

- E_n - Erkennbarkeit des nicht stimmhaften Explosivs
 F_0 - Grundton der Stimme
 F_{char} - charakteristisch verstärkte Frequenzbereiche
 T - Zeitfaktor (Dauer, Änderungsgeschwindigkeit...)

Eine Analyse des Artikulationsprozesses zeigt, daß nach einer relativ langen Unterbrechungsphase des Vokaltraktes eine jähe Öffnung folgt. Ihrer phasikalischen Realisierung entspricht ein Ablauf der nur einen Teil der mit $k_1(1 - \exp(-k_2t))$ ausgedrückten allgemeinen Exponentialkurve bildet. Das Öffnen des Stimmweges verläuft nur in erster Approximation sprunghaft, der Erregungsimpuls entsteht also mit beschränkter Geschwindigkeit. Das breiteste Spektrum erregt der Anlaufteil des Impulses, sofern dieser in genügender Intensität entsteht; dies wird zum größten Teil durch die Artikulationsweise beeinflusst. In der weiteren Phase des Öffnens entsteht ein homorganes postexplosives Reibegeräusch, modifiziert durch das Erweitern des Spaltes. Beide Geräuscharten werden durch die filtrierende Wirkung der Resonanzhöhlen beeinflusst. Auch die Parameter der Höhlen sind veränderlich, meistens zugleich mit Veränderungen des Geräusches. Folgt auf ein Explosiv ein weiteres Explosiv, endet der ganze Prozeß bei der dem zweiten Explosiv entsprechenden Okklusion. Das Geräuschsignal des ersten Explosivs ist so zeitlich begrenzt. Folgt ein Vokal, schließt sich verhältnismäßig bald nach der Explosion ein periodisches Signal an. Durch Resonanz wird dieses Signal in Frequenzbereichen verstärkt, die durch das ansonst abschwächende homorgane Geräusch des ausklingenden Explosivs betont werden. Das PE Geräusch wird kontinuierlich abgeschwächt und durch das periodische Signal maskiert; in seiner Anfangsphase kann also dieses zur Erkennbarkeit des Explosivs beitragen. Der letzte Teil des fließenden Überganges gehört zum typischen Bereich des folgenden Vokals. Beim Anhören des isolierten Teils aus dem mittleren Bereich des Überganges kann weder die Zugehörigkeit zum Vokal, noch diejenige zum Explosiv mit genügender Genauigkeit bestimmt werden.

Bei guter Aussprache sind im Spektrum zwei Teile zu erkennen: die Explosion und das postexplosives Geräusch, bei verschlechterter Qualität der Aussprache verschlechtern sich beide gleichzeitig. Bei geläufigem Sprechen ist die Explosion häufig geschwächt, oder (in spektraler Darstellung) fehlt sie ganz. Im Falle der segmentalen Darstellung des Sprechsignals (z.B. bei der LPC Synthese) kann die Explosion entweder ganz fehlen oder wird sie in der Länge des ganzen Segments dargestellt. Solche Weglassung oder Verlängerung wirkt sich meistens als störend aus.

EXPERIMENTELLE ANORDNUNG DER VERSUCHE

Die durchgeführten Experimente sollten folgende Fragen beantworten:

1. Welche ist die Bedeutung der eigentlichen Explosion und des PE Geräusches?
2. Welche sind die Folgen einer Verkürzung der eigentlichen Explosion bei fehlendem PE Geräusch?
3. Welche Folgen hat das Aussetzen der Explosion bei behaltemem PE Geräusch?
4. Welche sind die Folgen der fehlenden Explosion und des PE Geräusches bei behaltemem periodischen Anlaufteil des folgenden Vokals?
5. Welchen Einfluß auf die Perzeption des Explosivs hat das vollkommene Beseitigen des periodischen Anlaufteils des Vokals bei behaltener Explosion?
6. Gelten analogische Schlüsse, wenn das Explosiv vor einem anderen Explosiv oder einem Vokal steht?

Die Versuche wurden mit tschechischen Wörtern entsprechend den Formeln C_1C_2V und $C_1C_2C_3V$ durchgeführt. Untersucht wurde der Einfluß von Veränderungen am

- a) explosiv in der Initialstellung (C_1)
- b) explosiv in der Medialstellung (C_2 vor C_3)
- c) explosiv in der Medialstellung vor dem Vokal (C_2V)
- d) Übergangsbereich des Vokalanlaufs nach einem Explosiv (C_2V)

Die Worte wurden aus Tonbandaufnahmen guter Aussprache in einer akustisch gedämpften Kammer des Phonetischen Labors des Instituts für tschechische Sprache der ČSAV in Praha ausgewählt. Die Aufnahme wurde mit 2 männlichen Stimmen durchgeführt. Proben normaler Aussprache wurden nach den Regeln der zufälligen Anordnung mit mittels Montage modifizierten Proben vermischt, und ihre Perzeption wurde mittels einfacher Anhörteste überprüft. Bei der Probenverfertigung benutzte man ein Eichsignal 2 kHz, 100 ms, im Abstand 100 ms vor dem Probeanfang. In jedem Test wurden Proben derselben Stimme verglichen. Die Anhörgruppe betrug etwa 10 Personen (7+11); es handelte sich dabei um geborene Tschechen mit phonetischer Ausbildung sowie auch ohne dieser, jedenfalls mit normalem Gehör. Gewertet wurde nach einer dreistufigen Skala:

1. Normaler Gehöreindruck, ohne wahrnehmbare Veränderungen
2. Gehöreindruck mit wahrnehmbaren Qualitätsänderungen
3. Erkennbarkeit des Lautes gleich Null (der Laut nicht perzipiert, durch anderen Laut substituiert, nicht erkennbar).

Das Auswerten der Ergebnisse geschah nach Schema: Urteile (1+2) gegenüber 3.

Die Explosive nach a) und b) wurden in zwei Tests (A, B) verarbeitet. Die mittels Montage modifizierten Proben enthielten in diesen Fällen sowohl das Auslassen der Explosion (-E) als auch das Auslassen des homorganen PE Geräusches (-PE).

Die Explosive nach c) und der Übergang nach d) wurden in vier Tests verarbeitet (C, D, E, F). Im Test C wurden die Proben des normalen (V_n) oder veränderten (V_{med}) Vokals kombiniert mit Modifikationen des Explosivs mit ausgelassener Explosion (-E) oder ausgelassenem PE Geräusch (-PE). Die veränderte Probe (V_{med}) entstand dadurch, daß der normale Anlaufteil der periodischen Schwingung des Vokals durch den medialen - stationären - Teil des Vokals ersetzt wurde. Es entstanden 4 Modifikationen: $-EV_n$, $-PEV_n$, $-EV_{\text{med}}$, $-PEV_{\text{med}}$.

Der Test D enthielt Modifikationen der Länge des explosiven Geräusches vor dem Vokal mit ersetzttem Anlaufteil. Der Rest des PE Geräusches war 5 ms und die Explosion wurde auf die Dauer von 5, 10 und 15 ms begrenzt. Die Proben $E_{15+5V_{\text{med}}}$, $E_{10+5V_{\text{med}}}$, $E_{5+5V_{\text{med}}}$, $E_{0+5V_{\text{med}}}$.

Der Test E bestand aus Proben mit ausgelassener Explosion und ohne PE Geräusch, in denen entweder der normale Vokalanlauf folgte, oder war der Anlauf durch den Medialteil ersetzt. Ein langsamer Intensitätsanlauf dieses substituierten Teils erfolgte mit Längen von 10, 20, 40, 70 ms: so wurde das weiche Ansetzen der Intensität substituiert, bei dem die Komponente der Frequenzveränderung fehlte (die aber im normalen Sprechsignal immer vorhanden ist). So entstanden die Proben $-EV_n$, $-EV_{\text{med}10}$, $-EV_{\text{med}20}$, $-EV_{\text{med}40}$, $-EV_{\text{med}70}$.

Im Test F wurden Proben mit Explosion vor dem Vokal, dessen Anlaufteil durch den Medialteil ersetzt war, mit Proben ohne Explosion und PE Geräusch vor dem Normalvokal, oder vor dem ersetzten Anlaufteil kombiniert. Proben $+EV_{\text{med}}$, $-EV_n$, $-EV_{\text{med}}$.

Der angeführte Arbeitsgang ist anhand ausgewählter Beispiele dokumentiert. Die Abbildungen zeigen Spektrogramme des tschechischen Wortes mit Änderungen in Position nach c) und d) und Verarbeitung nach E, F. Es handelt sich um das Wort [ka:t] (=weben). Abb. 2 zeigt das Originalspektrum mit gut entwickeltem Übergang der Laute [ka:].

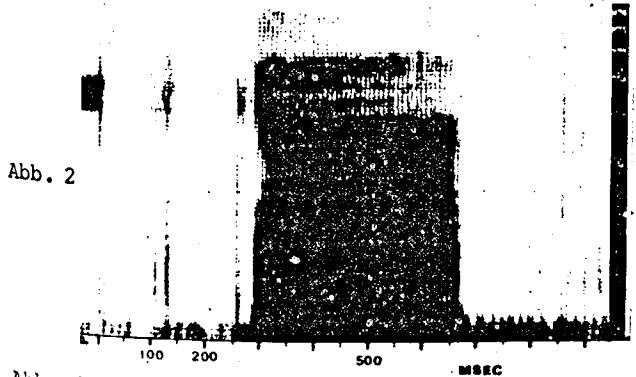


Abb. 2

Abb. 3 zeigt dasselbe Spektrum mit ausgelassener Explosion und ohne PE Geräusch.

Abb. 4 zeigt ein Spektrum mit behaltener Explosion und mit PE Geräusch. Der Anlauf-

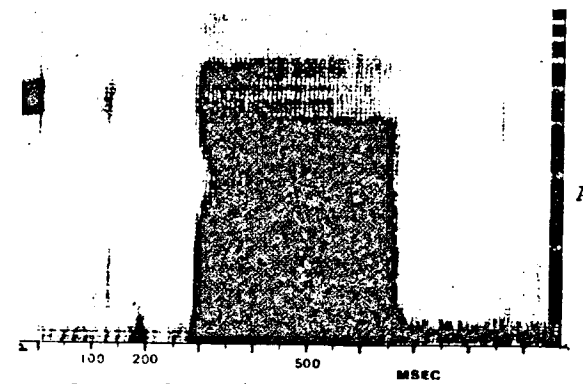


Abb. 3

teil ist durch Signal aus dem charakteristischen Medialbereich des folgenden Vokals [a:] ersetzt.

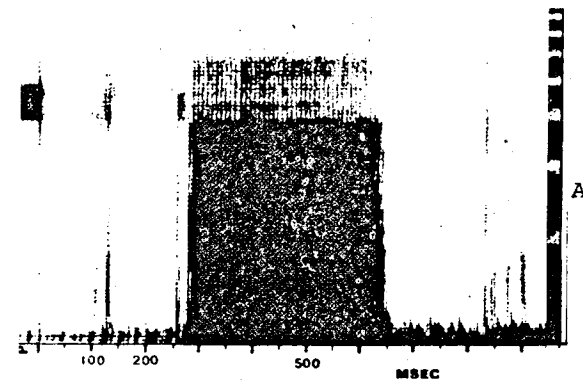


Abb. 4

Abb. 5 zeigt das Spektrum mit ausgelassener Explosion und PE Geräusch und mit ersetzttem Anlaufteil des Vokals (ähnlich wie in Abb. 4). Die Länge der Proben entspricht der Originalaufnahme. Bei der Auslassung des Originalsignals wurde ein Blank derselben Länge aus demselben Tonband benutzt. Partien ohne Sprechsignal sind mit gleichem Geräuschhintergrund der Tonaufnahme wie beim umgebenden Signal dargestellt. Die Montage nach Abb. 4 und 5 kann man künstlich (nur im Labor) erzielen. In der Sprache ist es infolge der endlichen Trägheit der sich bewegenden Stimmorgane nicht erreichbar. Zur Kontrolle diente die Anordnung nach Abb. 5, in der der Einfluß aller Signale des Explosivs vollkommen unterdrückt wurde. Ähnliche Verarbeitung wurde auch für den Fall mit dem Explosiv in der Medialstellung benutzt (C_2 in der Gruppe $C_1C_2C_3V$).

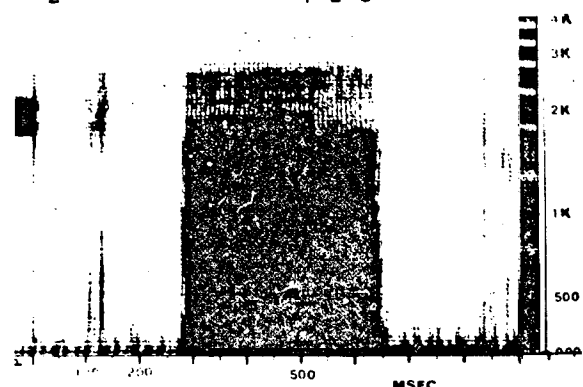


Abb. 5

Bei der Verarbeitung der Ergebnisse benutzte man statistisches Testverfahren. Jede Probe wurde im Test 2 + 3 mal präsentiert. In jedem Test wurde das arithmetische Mittel der richtigen Antworten und die Standardabweichung ermittelt. Richtige Antworten bewegten sich im Bereich ungefähr 83 + 100%.

AUSWERTEN DER ANHÖRTESTE

Das Auswerten der Teste ergab:

1. Die Perzeption des Explosivs in Medialstellung blieb im Prinzip ungestört:
 - a) bei Auslassen der Explosion, wenn ein PE Geräusch folgte,
 - b) bei Auslassen des PE Geräusches, wenn die Originalexpllosion vorhanden war.
2. Wurde das PE Geräusch ausgelassen, führte ein weiteres Kürzen der Explosion zu einer schlechteren Erkennbarkeit des untersuchten Explosivs. Im Grenzfall verschwand es vollkommen. Es handelte sich dabei um eine fehlende Perzeptionerscheinung, nicht um das Verwechseln mit einem anderen Laut.
3. Wurde im Signal nur das PE Geräusch behalten, blieb auch in diesem Falle die Perzeption weitgehend ungestört.
4. Wurden sowohl Explosion als auch PE Geräusch ausgelassen und der Anlaufteil des folgenden Vokals behalten (wie in Abb. 3), blieb die Perzeption ebenfalls ungestört.
5. Das Vertauschen des periodischen Anlaufteils des Vokals bei behaltener originaler Explosion (Abb. 4) hat auf die Explosivperzeption ebenfalls keinen bedeutenden Einfluß gehabt.
6. Analogische Schlußfolgerungen zeigen, daß jedes der drei untersuchten Elemente für die Perzeptionerscheinung allein genügt. Die Kombination von zwei oder allen Elementen hat redundanten Charakter und verbessert die Verlässlichkeit der Perzeption.

Im Kontrollversuch wurde die Explosion als auch das PE Geräusch ausgelassen und der periodische Anlaufteil des Vokals wurde durch den Medialteil ersetzt (Abb. 5). Das relevante Explosiv konnte keinesfalls identifiziert werden. Die ursprüngliche Gesamtqualität blieb dabei erhalten. Dies führte zu einer merklichen Verlängerung der Vokalperzeption. Bei Verlängerung des Intensitätsanlaufs der substituierten periodischen Komponente in Abb. 4 und 5 (schrittweise bis 80 ms) konnte die Voraussetzung nicht bestätigt werden, daß ein hartes (kürzeres) Ansetzen des periodischen Teils die Identifizierung des Explosives verbessert, sofern es nicht zugleich von einer entsprechenden Frequenzänderung begleitet wird.

Aus den Messungen folgt ebenfalls der Beweis der hohen Perzeptionsempfindlichkeit des Gehörs für signifikante Übergangsprozesse. Z.B. eine Verkürzung des relevanten Geräuschsignals des [k] bis auf ca 15 + 10 ms (Extremfall) hatte keine Einfluß auf die

Erkennbarkeit des Explosivs, unter der Voraussetzung einer Explosion mit gut entwickeltem Spektrum.

Im natürlichen Sprechsignal ergänzen sich also gegenseitig alle Komponenten: Explosion + PE Geräusch event. die periodische Komponente im Anlaufteil des folgenden Vokals. Bei idealer Artikulation ist das nach der Explosion folgende Signal redundant, umgekehrt bei schlechter Artikulation mit fehlender (ungenügend entwickelter) Explosion kann die zweite oder die dritte Komponente die Explosion vertreten. Besonders markant ist diese Tatsache bei der Kombination Explosiv + Vokal. Diese Wertung ist natürlich diskutabel: es kann ebenso behauptet werden, daß im geläufigen Sprechsignal der Einfluß der Übergangsgebiete primär zur Geltung kommt, während die eigentliche Explosion eher zur redundanten Information gehört.

SCHLUSSFOLGERUNG

Die beschriebene Methode demonstriert die Annahme, daß tschechische Explosive mittels mehrerer sich gegenseitig vertretender Merkmale bestimmt werden können. Die Spektrogramme der modifizierten Signale wurden zur Auswertung des Charakters von mittels der Perzeptionsteste identifizierten Proben benutzt.

Die Perzeptionsauswertung der tschechischen Explosivlaute beruht also auf der Bewertung der einander sich vertretenden Merkmale (Redundanzprinzip). Für richtige Bewertung des Explosives genügt nur ein einziges Merkmal von den hier beschriebenen (Einzelfälle natürlicher und künstlicher Sprachdegradation). Schlußfolgerungen kann man bei der Sprachsyntheseprogrammierung ausnutzen.

LITERATUR

1. F.S.Cooper, P.C.Delattre, A.M.Lieberman, J.M.Borst, L.J.Gerstman: Some experiments on the perception of synthetic speech sounds. JASA 24, 1952, 597-606.
2. E.Fischer Jörgensen: Tape cutting experiments with Danish stop consonants in initial position. ARIPUC 6, 1972, 104-168.
3. A. van Katwijk, J. t'Hart: Intelligibility of syllable - tied interrupted speech. I.P.O. Report April 1967, 99-102.
4. B.Lindblom: Accuracy and limitations of Sonagraph measurements. Proc. of the IV Int. Congr. of Phon. Sci., Helsinki 1961.
5. G.E.Peterson, W.S.Wang: Segmentation techniques in speech synthesis. JASA 30, 1958, 739-742.
6. G.Ungeheuer: Systematische Signaldestruktion als Methode der psychoakustischen Phonetik. Phonetica 18, 129.