

FÖRSÖK AVSEENDE VOKALTRANSITIONS RIKTNING
 OCH DESS BETYDELSE FÖR "PLATS"-DISTINKTIONEN
 BLAND TONANDE KLUSILER

Kurt Johansson

I en postumt utkommen uppsats i *Studia Linguistica* (1969) tar Delattre på nytt upp locusbegreppet till diskussion, och han påpekar härvid, att locus är fast och unikt för en given dialekt och vid en given grundtonsfrekvens, och att en sammanblandning med de konvergeringspunkter man erhåller vid mätningar på spektrogram av vanligt tal inte får ske. Konvergeringspunkterna, säger han, visar vad formanttransitions är, när de i mänskligt tal påverkas av koartikulation, locus däremot vad de måste vara för att bidra maximalt till perception av ett givet "artikulationsställe", både vad beträffar mänskligt och syntetiskt tal.

"The locus concept is a very economical one since all the different formant bends (transitions) that contribute to the perception of the same place of articulation can be specified by a single frequency, regardless of the vowel which precedes or follows the consonant." (Delattre, op.cit.).

Detta är uppenbarligen sant, men det ligger möjligen också en viss fara i användandet. Redan i tidigare Haskins-uppsatser (t.ex. hos Liberman et al., 1954) framhölls att "the direction and degree of second-formant transitions can serve as cues for the aurally perceived distinctions among the stop consonants", men i många senare arbeten tycks "degree" ha glömts bort och endast "direction" tagits i beaktande. Orsaken är troligen att söka i locus-begreppet och i påståenden om att transition måste peka mot en viss punkt, dvs. i en viss riktning, för att de skall ge intryck av att en viss konsonant, med ett visst artikulationsställe, producerats. "Accordingly, we propose to define the locus as the frequency toward which formant

transitions must point in order to contribute maximally to the perception of a given place of articulation." (Delattre, op.cit.).

Nedanstående serie av försök genomfördes som en följd av perceptuella undersökningar av svenska dentaler och supradentaler (vanligt tal), där det visade sig vara omöjligt att särskilja de båda konsonanttyperna enbart på grundval av vokal-(transitions-)informationen, efter det att endast en liten bit av de finala transitions¹ eliminerats. Detta var fallet inom samtliga berörda konsonantkategorier (tonlösa och tonande klusiler, nasaler, lateraler och frikativor). Den kvarvarande, största, delen av transitions borde utan tvivel ha haft tillräcklig längd för att man skulle ha kunnat uppfatta riktningen hos dem, om riktningen i sig självt skulle vara viktigaste "cue".

Författaren drog följaktligen den slutsatsen, att informationen i transitions riktning borde vara av sekundär betydelse, och transitions utsträckning, dvs. uppnåendet av vissa relativa frekvenser, det primära, när informationen från den egentliga konsonanten (när den nu är av intresse, vilket inte alltid är fallet) av en eller annan orsak inte når en lyssnare. Liknande tankegångar har framförts av Fischer-Jørgensen (1954).

För att testa hur det förhåller sig med övriga tonande klusiler inspelades följande stavelser: /a:b/, /a:d/, /a:g/, /ab/, /ad/ och /ag/. Talare var en manlig universitetsstuderande, vars tal kan sägas vara representativt för svenskt riksspråk. Varje stavelse upprepades tre gånger, och det mellersta exemplaret valdes till teststavelse.

I dessa stavelser gjordes sedan olika typer av ingrepp, dels manuellt, dels med hjälp av en elektronisk sax.

Stimuli presenterades sedan randomiserade för inalles 70 studenter via

1 De äkta supradentalerna kan i svenskan endast uppträda, om de föregås av vokal.

hörlurar av hög kvalitet. Lyssnarna kan förmodligen betraktas som "naiva", då det var deras första termin som fonetikstuderande, och eftersom de aldrig deltagit i test av detta slag förut.

Vid testtillfällena presenterades varje stimulus tre gånger i följd med en sekunds paus och återkom på samma sätt fem gånger per testavsnitt. De första tio stimuli tjänade endast till att göra lyssnarna bekanta med aktuell typ av stimuli, något som försökspersonerna själva inte var medvetna om. Efter var tionde stimulus gjordes en ca tio sekunder lång paus.

Lyssnarnas uppgift var alltså att avgöra om stimuli innehöll, eller hade innehållit, b, d eller g (forced choice)¹, och de förbereddes på att en del stimuli skulle komma att låta något onaturliga. Alla stavelser skulle betraktas som nonsens-stavelser.

Test IA

Vid dental-supradentalförsöken hade konsonant jämte finala transitions successivt eliminerats, varvid alltså borttagandet av ett relativt kort segment av transitions omöjliggjorde ett särskållande av dentaler och supradentaler.

I detta test var avsikten att undersåka om lyssnarna kan identifiera en tonande klusil med hänsyn till "place" enbart med hjälp av slutet på transitions, och, om detta är möjligt, hur mycket av transitions som behövs tas med.

Stimulustyp 1 utgjordes av oförändrade teststavelser. Typ 2 erhöåls däri- genom att explosioner jämte efterföljande schwa-segment klipptes bort (manuellt).

Övriga stimuli åstadkoms på så sätt att de initiala delarna av teststavelserna successivt skars bort med hjälp av den elektroniska saxen. För

1 I det sista testet stod valet i stållet mellan /a:/ och /a/.

att reducera antalet stimuli gjordes det första snittet (stimulustyp 3) ganska långt in i vokalen. Var snittet skulle göras, bestämdes av författaren auditivt. Borttaget segment förväntades inte kunna påverka konsonantuppfattningen särskilt mycket. Detta antagande tycks ha varit ganska realistiskt.

Stimulustyp 4 och följande innebar ett avkortande av teststavelserna ungefärligen i 20 ms-steg.

Alla stimuli, med undantag för den första typen, där explosionerna var kvar, bestod alltså av ett segment av vokalen följt av konsonantens "voice bar", vilken inte ansågs bära innehålla särskilt mycket platsinformation. Att något av konsonanten över huvud taget fick stå kvar, berodde på att jag om möjligt ville klargöra för lyssnarna, att det verkligen rörde sig om finala "unreleased" klusiler. I annat fall skulle återstående transitionsdelar kunna ha tolkats som isolerade explosioner, något som givetvis skulle ha varit intressant i sig självt (jfr test IIA och IIB). Ytterligare ett skäl var, att man sällan kan vara alldeles säker vid segmentering. Det finns ofta inte någon skarp gräns mellan det som brukar räknas till vokal respektive konsonant, och även om svårigheterna här inte är alltför stora, införs ett extra riskmoment redan från början. Vidare finns det knappast någon helt exakt klippmetod.

Om man emellertid har för avsikt att undersöka hur det förhåller sig med informationen i transitions, är det naturligtvis gynnsamt, om endast transitions är med. Detta var orsaken till att hela konsonanterna togs bort i test II och III.

Resultat

Inalles presenterades 63 stimuli 5 gånger för 20 lyssnare, vilket gav 6.300 lyssnarsvar.¹

När mer och mer av vokalen utesluts, sjunker värdena successivt för /a:b/, /ab/ och /a:d/. För övriga stavelser håller de högre värdena i sig längre, innan de faller snabbt.

Ett gradvis fall skulle kunna betraktas som ett indicium, att "place cues" verkligen är spridda över transitions. Ett visst mått av spridning är säkert vad man kan vänta sig, när vi närmar oss frekvenserna för transitions ändpunkter, även om riktningen i sig självt inte är det väsentliga, men det kan också ses som en naturligt konsekvens av att kortare, och därför "svårare", stimuli omväxlar med stimuli, där kanske hela vokalen finns med.

Klart är att sista delen (ca 20-25 ms) av transitions i regel bär den utan jämförelse mesta informationen. När enbart ändpunkterna² är med, är försökspersonerna i fem av sex fall säkra på sin sak.³ I synnerhet för /ag/, men även för övriga teststavelser, är det tydligt, att "the voice bars" inte innehållit så litet platsinformation som väntat. Vi får aldrig en jämn fördelning av lyssnarsvaren, när endast "the voice bars", eller delar av dem, återstår. Särskilt i den del som ligger närmast vokalen finns viss information.

-
- 1 Ibland saknas ett fåtal svar i de olika testavsnitten (dock ej här), varför angivna siffror i regel bör betraktas som ungefärliga.
 - 2 Med ändpunkter kommer i fortsättningen att avses de sista 20-25 ms av transitions.
 - 3 Skillnaderna mellan antalet rätta svar och antalet svar för den konsonant som favoriserats vid felsvar är i dessa fall signifikanta på 0,1 %-nivån ($p < 0,001$). För /a:d/ är denna skillnad däremot inte signifikant. 20 ms tidigare hade skillnaden legat på 0,1 %-nivån, och klipp i konsonanten resulterade omedelbart i b-svar ($p < 0,005$).

Test IB

Detta testavsnitt genomfördes i avsikt att undersöka om segment, som kan fungera i sin ursprungliga (finala) ställning, också räcker till i annan ställning i stavelsen, dvs. initialt.

Stimuli var i stort sett desamma som i föregående testavsnitt. Stimulustyp 1 (innehållande explosionerna) togs dock bort, liksom en del stimuli som inte bedömdes kunna påverka konsonantperceptionen särskilt mycket. På ställen som skulle kunna vara intressantare gjordes snitten tätare. Den viktigaste skillnaden mellan de båda testavsnitten var dock att stimuli i test IB spelades upp baklänges.

Resultat

Inalles presenterades 59 stimuli 5 gånger för 17 nya lyssnare, vilket gav 5.015 bedömningar.

Redan från början (stimulustyp 2) blir antalet rätta svar stundom avsevärt lägre än för motsvarande stimulustyp i föregående test. Men endast /a:d/ ger ett övervägande antal felaktiga svar (ej signifikant skillnad). De lägre värdena är naturligtvis vad man kan vänta sig, när stimuli spelas upp baklänges.

På det hela taget kan man säga, att ursprungligt finala segment kan fungera också i initial ställning. Ingenting motsäger tidigare påstående, att transitions ändpunkter skulle vara av större betydelse än transitions riktning. Möjligen d-data i så fall, men svaren för båda d-stavelserna är så osäkra, att de inte kan ligga till grund för slutsatser. d-stimuli låter ungefär som om de börjar med ett ljud snarlikt den engelska tonande frikativan [ð], och mera fonetiskt bevandrade lyssnare hade förmodligen inte haft några svårigheter med identifikationen.

Eftersom "the voice bars" inte har visat sig vara helt neutrala och tycks

innehålla mer information i initial än i sin ursprungliga ställning, är det uppenbarligen svårt att avgöra hur mycket av transitions som måste tas med för erhållande av rätta svar.

Test IIA och IIB

I dessa testavsnitt användes endast stavelserna med korta vokaler. För att erhålla skarpa gränser, i band såväl som i spektrogram, till att användas som utgångspunkt för mätningar, klipptes en mycket kort bit i början av vokalerna bort. Alla ingrepp gjordes här, liksom nedan, manuellt.

Eftersom "the voice bars" uppenbarligen innehåller en del platsinformation, var nästa steg att klippa bort även dessa. Avståndet från de ovan nämnda sonagramgränserna till ingreppspunkterna översattes till bandhastighet, varpå konsonanterna togs bort. Metoden visade sig fungera fullt tillfredsställande.

De på detta sätt producerade "stavelserna" utgjorde stimulustyp 1 (S1).

Transitions är utan tvivel viktigast för perception av "place", av det som nu återstår. Längden hos vokaltransitions är visserligen något olika i de här använda stavelserna¹ men nedan har jag endast laborerat med de sista 80 ms av varje vokal.

S2 erhöles genom att de sista 20 ms togs bort. I S3 kvarstod endast denna del av de sista 80 ms. Resten hade ersatts med "tomt" band. Bandet var i själva verket inte alldeles tomt, då bakgrundsbruset hållits på samma nivå som för inspelade stimuli.

Beträffande övriga stimuli hänvisas till fig. 2, där de större rektangelarna motsvarar de tidigare nämnda 80 ms-avsnitten. Det bör också framhållas, att vokalernas "steady states" också ingick i stimuli, men att de bibehölls oförändrade.

¹ Ungefärligen 100 ms för /ab/, 105 ms för /ad/ och 110 ms för /ag/. Mätningarna är gjorda på längsta formantböjning.

Stimuli arrangerades sedan på samma sätt som i föregående testavsnitt. Skillnaden mellan IIA och IIB bestod i att stimuli i IIB spelades upp baklänges.

Resultat

I vardera testavsnittet presenterades 21 stimuli 5 gånger för 18 nya lyssnare, vilket gav 3.780 bedömningar.

Även om "the voice bars" inte är utan information, kan de inte betraktas som primära "cues". I allmänhet tycks lyssnarna inte ha några verkliga svårigheter att tolka rätt, inte ens när hela konsonanten är borta (S1 i fig. 2), men liksom i föregående reverstest utgör /d/ ett problem, och i stället tolkas i regel ett /b/ in. För en jämförelse med de föregående testavsnitten, där "the voice bars" är medtagna, se fig. 1.



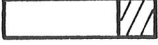




	/ab/	/ab/ _r	/ad/	/ad/ _r	/ag/	/ag/ _r
Stimuli inkluderande "voice bars"	100 %	78 %	98 %	60 %	98 %	73 %
Stimuli utan "voice bars" (S1)	85 %	88 %	63 %	41 % ¹	70 %	50 %

Fig. 1. Procent rätta svar för stimulustyp 1 (S1) och för motsvarande stimuli i tidigare test, där "the voice bars" inkluderats. Reversstimuli har markerats med r-index.

Ett borttagande av transitions ändpunkter (S2) påverkar inte bedömningen av b-stimuli särskilt mycket (se fig. 2). För /d/ är osäkerheten betydligt större, men något förvånande är, att d-svaren är fler för baklängesspelat

1 Ej signifikant skillnad.

STIMULUS

		b		d		g	
S1		b	85	88	28	50	7 28
		d	9	8	63	41	23 22
		g	6	4	9	9	70 50
S2		b	82	74	47	42	19 42
		d	13	22	51	50	40 23
		g	4	3	2	8	41 34
S3		b	64	58	43	22	18 26
		d	19	29	37	49	26 22
		g	17	13	20	29	57 52
S4		b	82	70	31	29	9 22
		d	7	14	48	43	27 10
		g	11	16	21	28	64 68
S5		b	81	69	30	42	2 16
		d	11	17	53	36	34 11
		g	8	14	17	22	63 73
S6		b	63	54	19	23	12 19
		d	17	27	50	56	30 21
		g	20	19	31	21	58 60
S7		b	72	68	21	26	8 22
		d	13	22	60	63	33 23
		g	14	10	19	11	59 54

RESPONSE

Fig. 2. Fördelning av lyssnarsvaren (i procent) för test IIA och IIB. Till vänster (S1 etc.) typ av ingrepp. Av siffrorna i de olika rutorna avser första kolumnen test IIA, andra test IIB.

/ad/, än de var, när de sista 20 ms inkluderats, men skillnaderna mellan antal d-svar och antal svar för den kategori försökspersonerna främst favoriserat vid felbedömning är inte i någondera fallet signifikanta. Rättvänt */ag/* ger gissningsvärden, och även om samma stimulus i reverstestet ger övervägande b-svar är skillnaden inte heller här signifikant.

Klipps de första 60 ms bort, så att endast de sista 20 ms återstår (S3), får vi för */d/* (rättvänt) 6 % fler b- än d-svar (ej signifikant). I övriga fall står det fullt klart, att lyssnarna inte gissar, inte ens för */d/* (revers), där $p < 0,05$. d-siffran är här av samma storleksordning som för S2, men övriga svar är här jämnare fördelade. För */g/*, där F2 och F3 konvergerar i det sista segmentet, blir värdena i båda testavsnitten betydligt högre än för S2 ($p < 0,005$ i båda fallen). Både vad beträffar */d/* och */g/* är reversvärdena högre än för S1, men även */b/* ger höga värden ($p < 0,0005$ för */ab/* och $0,005$ för */ab/*_r). - I samband med */b/* är transitions för F2 och F3 mindre påfallande än för de andra "teststavelserna", och det borde alltså vara av mindre betydelse, om en bit av dem tas bort. I vilket fall som helst är b-värdena högre för S2 än för S3.

Både S2 och S3 för */d/* (rättvänt) förorsakar stor osäkerhet. Denna osäkerhet uppträdde också i de tidigare testavsnitten.

I regel kan alltså sista delen av transitions fungera som "place cue", utan hjälp av intilliggande konsonant, och trots att stimuli inte låter fullt naturliga med ett 60 ms långt tomt segment mellan vokalernas "steady states" och transitions ändpunkter och trots avsaknad av normala transitions. Speciellt reversstimuli borde, tycker man, förorsaka svårigheter. I IIA skulle ändpunkterna möjligen kunna tolkas som explosioner, då vi i själva verket genom våra ingrepp i bandet åstadkommit nya konstgjorda klusiler med tonlösa ocklusioner och tonande explosioner. I reverstestet kom

ju explosionerna före ocklusionerna.¹ "Explosionerna" har ungefärligen den duration som använts vid syntesförsök (se t. ex. Copper et al., 1962). Vad ocklusionslängden och dess betydelse för perceptionen anbelangar är det svårt att spåra någon definitiv trend. För att få klarhet på denna punkt måste speciella försök göras. I ursprungsstavelserna varierar dock denna duration något från en stavelse till en annan.

I IIA och IIB manipulerades transitions i 20 ms-steg. Detta är troligen en förenkling. Förmodligen är de viktiga segmenten av något olika längd för olika "artikulationsställen" för vokal och konsonant. Jämför vi S3 och S4 finner vi att värdena, med ett mindre undantag, ökar för S4, men eftersom transitions- och explosionsduration varierar samtidigt, uppstår viss osäkerhet beträffande tolkningen av resultaten.

Endast för /d/ (revers) är värdet något lägre för S4 än för S3 och en övergång till icke signifikant skillnad inträffar. Det mest påtagliga är annars förändringen till relativ säkerhet för /d/ (rättvänt) ($p < 0,10$).

S5 innebär, med undantag för /d/ (revers), där b-svaren än en gång är något i övervikt (6 %), inte någon större förändring jämfört med S4.

S6, innehållande det första och det sista 20 ms-segmentet, ger med undantag för /d/ ungefär samma värden som S3. Försökspersonerna gissar alltså inte, och vi kan med fog anta, att de första 20 ms inte är av särskilt stor betydelse.

Tar vi med de första 40 ms (S7), låter stimuli något naturligare, något som slår igenom i lyssnarsvaren i de flesta fall. Jämför vi med S5, där "ocklusionen" är av samma längd, i stället för med S6, finner vi att S7-värdet stiger för /d/, faller eller är ungefärligen oförändrat för de andra konsonanterna. Samma är förhållandet vid jämförelse med S4. För alla tre

1. Men här är att märka att försökspersonerna var förberedda på att stimuli innehållande /b/, /d/ eller /g/ skulle presenteras. Annars föreligger uppenbarligen en tendens att under vissa förutsättningar perceptuellt tidigarelägga en akustisk "störning" (Ladefoged, 1967).

konsonanterna tycks det näst sista segmentet innehålla viss information, men den allra sista delen förefaller vara klart viktigast.

I S7 är d-värdena höga i jämförelse med de flesta av övriga d-stimuli, och följaktligen skulle de första delarna av transitions vara av större betydelse här, men man kan inte bygga alltför mycket på så oregelbundna resultat. Viss information finns även i motsvarande segment för övriga konsonanter, men den är mycket liten.

Som sammanfattning av dessa testavsnitt kan alltså sägas, att än en gång har den allra sista biten av transitions visat sig innehålla kraftiga "place cues". De sista 20 ms räcker i de flesta fall väl för ett särskällande av /b/, /d/ och /g/, men försöken visar, att även det näst sista segmentet är av viss, men knappast avgörande, betydelse.

Vidare har det också här visat sig, att ursprungligt finala segment även förmår att fungera i initial ställning.

Test IIIA och IIIB

Utgångspunkt för dessa testavsnitt var det, vid tillfället för testförberedelserna, något vågade antagandet, att det allra sista 20 ms-segmentet av vokaltransitions skulle innehålla "place cues". För erhållande av stimuli 8-13 skars detta segment bort (manuellt) och ersattes med motsvarande segment från andra "stavelser". Se fig. 3 där även stimuli 1-3 från test IIA respektive IIB tagits med för jämförelses skull.

Stimuli avlyssnades vid samma tillfälle som dessa båda testavsnitt. IIA och IIIA presenterades tillsammans, med samtliga stimuli randomiserade, och sak samma gällde för IIB och IIIB.

På detta sätt vägdes alltså informationen i sista segmentet i en "stavelse" mot de första delarna av transitions i en annan.

Det förtjänar att nämnas, att stimuli givetvis innehåller hörbara transitioner, när övergången mellan höga och låga frekvenser blir alltför

abrupt, men de var på intet sätt slående i jämförelse med vad som förekom i stimuli ingående i test II, och inte ens för dessa var reaktionen särskilt "negativ" beroende på "onaturliga" stimuli.¹ Efter testen framförde en del lyssnare, att de hade hört något slag av oidentifierbar "störning", men något konsekvent utfrågande på denna punkt genomfördes inte.

Resultat

De här beskrivna testavsnitten, vilka alltså presenterades tillsammans med motsvarande avsnitt från test II, omfattade vardera endast 6 stimuli, vilka avlyssnades 5 gånger av 18 försökspersoner. Vid sidan om de 3.780 bedömningar från test II erhöles alltså ytterligare 1.080.

I fig. 3, där alltså stimuli 1-3 från föregående test tagits med, visas typen av förändring i stimuli.

Stimuli 1-3 har tidigare kommenterats.

Stimuli 8 och 9. d- och g-ändpunkterna förmår inte förändra svaren från b. Detta gäller båda stimuli, och är i sig inte särskilt förvånande. Ett kort segment ställs mot en kompakt, ganska oföränderlig energikoncentration. Inte desto mindre sänker g-ändpunkterna b-svaren i IIIB med 50 % jämfört med stimulus 1 ($p < 0,10$). I alla övriga fall är b-värdena slående höga ($p < 0,0005$). Detta är i synnerhet förhållandet för stimuli innehållande d-ändpunkter. Det högre värdet för stimulus 8 (rättvänt) i jämförelse med motsvarande stimulus 1 kan förklaras med det faktum att även stimulus 3, där endast d-ändpunkterna är kvar, får lyssnarna att i viss mån favorisera b-svar.

Stimuli 10 och 11. b-ändpunkterna åstadkommer en ansevärd förhöjning av antalet b-svar, om vi jämför med stimulus 1, men för stimulus 10 (IIIA)

1 Här är också att märka att försökspersonerna väntade sig klusiler, dvs. ljud som normalt innehåller transienter.

STIMULUS

RESPONSE
IIIA IIB

		b	d	g	b	d	g
1		86	9	6	88	8	4
2		82	13	4	74	22	3
3		43	37	20	22	49	29
3		18	25	57	26	22	52
8		96	2	2	74	21	4
9		78	13	9	44	28	28

STIMULUS

RESPONSE
IIIA IIB

		b	d	g	b	d	g
1		28	63	9	50	41	9
2		47	51	2	42	50	8
3		64	19	17	58	29	13
3		18	25	57	26	22	52
10		42	48	10	69	22	9
11		22	34	43	34	28	38

STIMULUS

RESPONSE
IIIA IIB

		b	d	g	b	d	g
1		7	23	70	28	22	50
2		19	40	41	42	23	34
3		64	19	17	58	29	13
3		43	37	20	22	49	29
12		6	49	46	37	14	49
13		8	39	53	29	22	49

- b transitions
- d ---
- g ---
- empty tape

Fig. 3. Fördelning av lyssnarsvaren (i procent) för test IIIA och IIIB.

har vi fortfarande ett fåtal fler d-svar (men ej signifikant skillnad), dvs. ungefär samma som för stimulus 2. I IIIB är det emellertid en klar dominans för b ($p < 0,0005$). Detta är givetvis också beroende på det höga antalet b-svar redan för stimulus 2. g-ändpunkterna (S11) förorsakar övervikt för g-svar, men de första delarna av transitions syns inte vara helt utan information, eftersom denna övervikt inte är signifikant.

Stimuli 12 och 13. Stimulus 12 (IIIA), där första delen av g-transitions kombinerats med b-ändpunkter, ger d- och g-svar (49 % resp. 46 %) och de höga värdena kan kanske synas vara ett indicium för resultaten från de ursprungliga Haskinsexperimenten med F2 (Delattre et al., 1955), nämligen att g karakteriseras av ett högt, b av ett lågt och d av ett mellanliggande locus. Ser vi emellertid på motsvarande stimulus 2 finner vi att svaren är ungefär desamma här. b-ändpunkterna har i själva verket åstadkommit mycket litet, trots det höga värdet för stimulus 3. Stimulus 12 (IIIB) medför övervägande g-svar, trots vad man kunde vänta sig med utgångspunkt från stimulus 2 och 3. Skillnaden är dock inte signifikant. d-ändpunkterna i stimulus 13 ändrar lyssnarsvaren i riktning från b-svar till g-svar. Endast i IIIB är skillnaden signifikant ($p < 0,05$). Lyssnarna blir alltså säkrare på att stimulus inte innehållit /b/.

Här har än en gång visat sig att g-ändpunkterna har avsevärt lättare att göra sig gällande än motsvarande segment i samband med de andra konsonanterna. Förmodligen skulle det sista segmentet ha gjorts något längre i dessa fall, om de skulle ha haft framgång i att överflygla g-transitions första delar.

Ingendera delen av d-transitions tycks innehålla så mycket information som motsvarande partier för övriga konsonanter, något som bestyrker tidigare iakttagelser.

När stimuli spelas upp baklänges, visar lyssnarna en ökad tendens att tolka i enlighet med ändpunkterna. I 4 av 6 fall blir antalet rätta svar - dvs. rätta med utgångspunkt från ändpunkterna - högre i reverstestet. Värdena stiger med mellan 17 % och 31 % i dessa 4 fall, medan de i de båda andra fallen (stim. 11 och 13) är 5 % respektive 17 % lägre.

Det är alltså möjligt, att ett segments position i viss mån kan påverka svaren.

Betraktar vi de svar som rätta, vilka avgivits i enlighet med första delen av transitions, finner vi, att vi i 5 av 6 fall får lägre värden, när stimuli spelas baklänges, och delarna i fråga förekommer efter ändpunkterna, men speciellt för /g/, där det sjätte fallet ingår, är skillnaderna mycket små. Här är dock att beakta, att den första delen av transitions alltid har något segment före sig, och att den i de rättvända stimuli m.a.o. endast kommer först av de segment som främst skulle kunna påverka konsonantperceptionen.

Test IV

/a:/ och /a/ har i svenskt riksspråk klart skilda vokalkvaliteter¹, vilket givetvis också medför olika vokaltransitions. Vi har alltså två typer av riktningsskillnader i de teststavelser vi här använt, dels den skillnad som inträffar hos /a:/:s (eller /a/:s) transitions i samband med olika artikulationsställen för konsonanterna, dels den olikhet som inträffar när /a:/ respektive /a/ kombineras med samma konsonant.

Vid förberedelser för tidigare försök tyckte jag mig uppfatta viss skillnad i vokalkvalitet i det senare fallet, trots att ganska långa delar av transitions tagits bort, och endast kortare finala segment var kvar. Detta

1 Fonetiskt: [ɑ] resp. [a].

är väl vad man kan vänta sig, men frågan är då om "naiva" lyssnare kan uppfatta denna skillnad, och om så är fallet, hur mycket av transitions som måste vara med.

Om ett särskiljande är möjligt, skulle riktningen alltså kunna träda in som "cue" på ett annat område.

Stimuli var till stor del desamma som i I B, fast de här användes rättvända.

Som ena ytterlighet presenterades hela vokalerna tillsammans med konsonanternas "voice bars". Explosionerna hade tagits bort, eftersom de i viss grad skulle kunna (?) reflektera kvaliteten hos föregående vokal.

Den andra ytterligheten var, att endast delar av "the voice bars" presenterades, något som rimligtvis borde medföra gissning. Mellanliggande stimuli bestod av kontinuerligt avkortade vokaltransitions jämte "voice bars".

Stimuli omfattade maximalt vokalsegment av nedanstående duration¹:

/a:b/	55 ms	/a:d/	75 ms	/a:g/	65 ms
/ab/	60 ms	/ad/	65 ms	/ag/	65 ms

Som synes är dessa vokalsegment inte särskilt långa, men de är ungefär dubbelt så långa som den duration en vokal måste ha för att över huvud taget kunna tolkas som vokal (Delattre, 1965). Men undantag för /a:d/ är minst hela transitions med i samband med de långa vokalerna.

Resultat

I detta testavsnitt presenterades 55 stimuli 5 gånger för 15 nya lyssnare, vilket gav 4.125 bedömningar.

Svaren för fyra typer av stimuli ges i fig. 4.

1 Undantag: stimulustyp 1, där hela vokalen var med.

	/a:b/	/ab/	/a:d/	/ad/	/a:g/	/ag/
A	100 %	92 %	100 %	92 %	100 %	91 %
B	49 %	91 %	41 %	89 %	51 %	82 %
C	36 %	86 %	25 %	77 %	37 %	61 %
D	47 %	53 %	29 %	60 %	44 %	71(51) ¹ %

Fig. 4. Procent rätta svar för en del av stimuli. A innebär att vokaler-
na har ursprunglig längd, B att hela eller större delen av tran-
sitions fortfarande finns kvar, C att ca 20 ms av dessa återstår,
och D avser värdena ungefär vid VC-gränsen.

När hela vokalen var med (A i fig. 4), erhöles 100 % rätta svar för stimuli innehållande lång vokal. Varför stimuli med kort vokal endast ger 91-92 % är oklart.

För D, dvs. ungefär vid VC-gränsen, överväger /a/-svaren², likaså för C,³ där ca 20 ms av vokaltransitions medtagits. I det senare fallet ökar faktiskt /a/-svaren i samtliga fall med undantag för /ag/, som med sina två värden för D är tveksamt.

När hela eller största delen av transitions ingår i stimuli (B), ökar säkerheten i samtliga fall, men inte i någon imponerande grad.⁴ För den långa vokalen har vi alltså endast uppnått ren gissning, vilket är för-
vånande med tanke på att vi i två fall är inne i vokalens "steady state".

-
- 1 Att två siffror tas med för denna stimulus beror på att VC-gränsen går ungefär mitt emellan två snitt.
 - 2 De värden som inte representerar gissning (/a:d/ och /ag/ [första vär-
det]) avviker mer än 3σ från det teoretiska medelvärdet.
 - 3 /a:d/, /ab/ och /ad/ avviker mer än 3σ , /a:b/ och /a:g/ mer än 2σ .
 - 4 För den korta vokalen är avvikelserna i samtliga fall mer än 3σ , för den långa får vi ren gissning.

Längden tycks alltså vara betydligt intressantare för vår perception än vokalkvaliteten medan transitions på sin höjd kan betraktas som sekundära "cues".

Sammanfattning

Ovan har 7 deltest redovisats, där inalles ca 20.300 lyssnarsvar avgivits av 70 lyssnare på 231 olika stimuli.

Test IA genomfördes i avsikt att utröna om sista delen av transitions (tillsammans med klusilernas "voice bars") kan ge tillräckligt information beträffande konsonantens "place of articulation". Trots att "the voice bars" inte var så fria från information av detta slag som väntat, måste svaret bli jakande. Försökspersonerna behövde i regel ca 20-25 ms av den allra sista delen av transitions för att avge övervägande korrekta svar.

Test IB visade, att segment som kan fungera i final ställning också i regel förmår ge tillräcklig information i initial ställning. Eftersom "the voice bars" inte var helt "färglösa", och i själva verket tycktes bära mer information i initial än i final ställning, var det här inte möjligt att avgöra hur mycket av transitions som behöver ingå i stimuli för erhållande av rätta svar.

Test IIA och IIB, där även "voice bars" eliminerats, bekräftade att de sista 20 ms av transitions i regel mycket väl kan innehålla tillräckliga "place cues", och vidare att dessa segment kan fungera såväl i final som i initial ställning. Även det näst sista 20 ms-segmentet innehåller viss, men oftast inte avgörande, information.

Samtliga dessa experiment, och även IIIA och IIIB, visar, att de olika delarna av d-transitions inte innehåller lika starka "cues" som motsvarande segment av b- och g-transitions. Åtminstone i samband med de här använda vokalerna har g-ändpunkterna lättast att göra sig gällande.

Riktningen för hela transitions är alltså, av ovanstående experiment att döma, av sekundär betydelse, och informationen i sista delen av transitions det primära.

Test IV genomfördes för att se om transitions riktning möjligen kan ge information av annat slag, nämligen beträffande vokalkvaliteten i stället. De vokaler, som använts, ligger visserligen ganska nära varandra ur denna synpunkt, men de ger dock så olika transitions, att detta mycket väl vore tänkbart. Det framgick emellertid, att dessa även här är av underordnad betydelse. Man kan knappast tolka resultaten på annat sätt än att vokalduration är viktigare än vokalkvalitet för distinktion mellan här berörda vokaler (jfr resultaten för andra vokaldistinktioner hos Hadding-Koch och Abramson, 1966).

Diskussion

Transitions riktning, såsom den definieras genom locus-begreppet, har följaktligen visat sig vara en sekundär "cue", och den viktigaste informationen ligger alltså i den allra sista delen av transitions.

Frågan är då: Är de delar av transitions, som visat sig tillräckliga för en riktig identifikation inom kategorin tonande klusiler, trots allt av tillräcklig längd för att möjliggöra ett uppfattande av frekvensförändring inom segmenten?

Andra perceptionsförsök talar mot en sådan tolkning. För att kunna ge intryck av tonhöjd måste en ton ha en approximativ längd av 10-20 ms, och dessa värden förefaller mycket kritiska även vid andra typer av avlyssningstest. Se exv. Cherry (1966, sid. 130).

I ovanstående experiment rör vi oss med segment av ungefär denna duration, men här skulle det samtidigt vara nödvändigt att uppfatta en förändring, och riktningen hos denna förändring, något som torde vara en alldeles för svår uppgift, särskilt som F2 och F3 ibland går åt samma håll, ibland åt olika (se närmare nedan).

Heinz et al. redovisade 1967 undersökningar där försökspersoner fått matcha tonhöjden hos en jämn ton med en ton med hastigt förändrad frekvens (med en duration på 20 eller 50 ms). Lyssnarna hade inte några svårigheter att tillskriva den sistnämnda tonen en viss tonhöjd, och till yttermera visso låg bedömningsfrekvenserna just i närheten av slutfrekvensen. Liknande försök med likartade resultat presenterades 1961 av Brady et al. Dessa båda försök stöder alltså resultaten från de här genomförda experimenten.

Formantbøjningarnas längd har inte visat sig vara av fundamental betydelse för distinktioner inom samma konsonantkategori (Lieberman et al., op.cit.). Test IA, där vokalernas "steady states" tillsammans med större delen av transitions seknades, bekräftar detta. Inte heller kan formantfrekvenserna i "steady states" vara av primärt intresse.

Med utgångspunkt från explosionerna har det ofta framhållits, att /b/, /d/ och /g/ kännetecknas av energikoncentrationer vid låga, höga respektive mellanliggande frekvenser, vilket skulle motivera en uppställning enligt distinktiva drag som gjorts för såväl vokaler som konsonanter av Jakobson och Halle (1956):

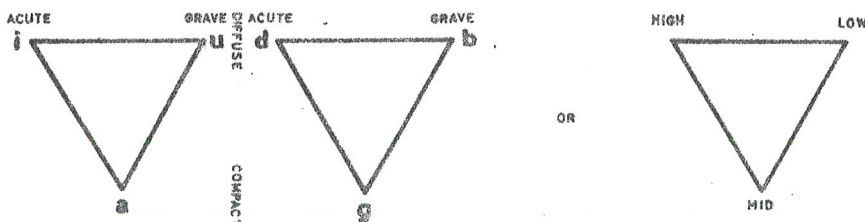


Fig. 5. Jakobson - Halle - systemet.

Ett studium av energikoncentrationerna vid VC-gränserna i våra teststavelser motsäger inte en sådan uppställning¹. Också andra undersökningar

¹ Detta gäller också data rörande de egentliga explosionerna.

tycks bekräfta den. Lisker et al. (1960) visar bl.a. de syntetiska stimuli som ger högst lyssnarvärden för /ba/, /da/ och /ga/, där detta klart framgår. Här hade de lägsta fyra formanterna inkluderats. Den raka fjärdeformanten var densamma i de tre stavelserna. En förskjutning uppåt enbart av tredjeformantens transition var tillräcklig för en förändring från g- till d-svar, och på det hela taget tycks denna formant ha visat sig besitta en större distinktiv roll än fallet hade varit vid motsvarande Haskinsexperiment (Harris et al., 1958).

Syntesförsök (vid Haskinslaboratorierna) med endast två vokalformanter hade resulterat i ett högre locus, om något, (Delattre et al., op.cit.), för /g/ än för /d/, vilket förefaller otillfredsställande med tanke på tre- och fyrformantstudierna.

När högre former saknas, borde andraformanternas böjningar betraktas som förändringar i något slag av effektiv andraformant, dvs. som transitions hos andraformanten och tredjeformanten i naturliga stavelser. Se t.ex. Fischer-Jørgensen (1958). Om man emellertid betraktar Haskinsexperimenten närmare, finner man att andraformantens transitions för främre, eller orundade, vokaler slutar högre upp i frekvensskalan för /g/ än för /d/, och följaktligen skulle Jakobsons uppställning passa dåligt här.

En alternativ tolkning skulle kunna vara, att /b/ och /g/ i en given omgivning karakteriseras av energikoncentrationer vid mera definitiva frekvenser medan /d/ stundom har en dominans för höga frekvenser, stundom ett spritt eller neutralt spektrum. Detta är i överensstämmelse med vad Halle et al. (1957) funnit beträffande de egentliga konsonantsegmenten: "/t/ and /d/, the postdental stops, have either a flat spectrum or one in which the higher frequencies (above 4000 cps) predominate, aside from an energy concentration in the region of 500 cps".

I de olika testen har dentalkonsonanten varit svårast att tolka rätt, och valet har alltså i stor utsträckning stått mellan /b/ och /g/. Svaren

tycks alltså bekräfta det mera definitiva för dessa konsonanter, men de skulle likaväl kunna vara indicium på det som framgick ur Haskins-försöken, dvs. att andra formantens transitions för främre (eller orundade) vokaler tillsammans med /g/ tycks peka mot ett högre locus, och att /b/ och /g/ följaktligen också ur denna synpunkt skulle utgöra ytterligheter. Detta motsägs bl.a. av sådana försök där även F3 tagits med. I vilket fall tycks /d/ vara diffusare, i ordets egentliga mening.

Oavsett vilket, vore det tillfredsställande, om olika segment, i de fall de innehåller drag avseende samma sak¹, verkligen pekade i samma riktning, dvs. om relationen mellan manifestationer av olika fonem inom samma konsonantkategori alltid är densamma i en given ställning, vare sig data erhålls från den egentliga konsonanten eller närliggande segment.

Detta är, som jag ser det, en nödvändig förutsättning för Fant's antagande (1967) att vi rör oss med någon sorts integrerat värde för flera cues.

Koartikulation för en konsonant med såväl föregående som efterföljande vokaler skulle uppenbarligen fordra en anpassning av ändpunkterna för den första vokalens transitions också till följande vokal (Öhman, 1966), men detta skulle inte i en given omgivning innebära någon förändring, om ovanstående antagande är riktigt.

Vi har alltså till följd av de här genomförda försöken delvis återvänt till en ursprungligare åsikt beträffande vokalernas transitions, dvs. att de till största delen endast tycks vara ofrånkomliga resultat av förändringen i artikulation vid övergången mellan vokal och konsonant, med det viktiga tillägget, att den del av vokaltransitions, som ligger närmast konsonanten i regel innehåller starka "place cues" beträffande konsonanten i fråga. Delattres tes (1963) "Consonant is movement; vowel is steady state," tycks m.a.o. inte vara helt tillämplig.

1 Som framhålls av Delattre (1969) kan informationen i explosioner och transitions vara olika, när velara klusiler förekommer tillsammans med rundade vokaler.

Summary

In an earlier study of Swedish dental and supradental (often retroflex) consonants appearing after different vowels the author found that the elimination of a rather short final segment of the final vowel transitions made a distinction regarding the "place of production" of the consonants impossible.

In the experiments reported above the information-carrying capacity of the corresponding portions of vowel transitions appearing together with /b/, /d/, and /g/ has been tested, and it seems quite clear that as a rule these last segments contain the major part of the information. This suggests that the direction of the vowel transitions is not the essential cue for a distinction according to "place of production" of a consonant.

The investigations will be published in a more extensive form in Travaux de l'institut de phonétique de Lund.

Litteratur

- BRADY, HOUSE, STEVENS, Perception of sounds characterized by a rapidly changing resonant frequency, Journ. Acoust. Soc. Am., 33:10, s. 1357-1362, 1961
- CHERRY, C., On human communication, 2nd edition, The M.I.T. Press, 1966
- COOPER, DELATTRE, LIBERMAN, BORST, GERSTMAN, Some experiments on the perception of synthetic speech sounds, Journ. Acoust. Soc. Am., 24:6, s. 597-606, 1952
- DELATTRE, P., Change as a correlate of the vowel-consonant distinction, Studia Linguistica, XVIII:I, s. 12-25, 1964
- Coarticulation and the locus theory, Studia Linguistica, XXIII:I, s. 1-26, 1969
- Research techniques for phonetic comparison of languages, IRAL, I/2, s. 85-97, 1963
- DELATTRE, LIBERMAN, COOPER, Acoustic loci and transitional cues for consonants, Journ. Acoust. Soc. Am., 27:4, s. 769-773, 1955
- FANT, G., Auditory patterns of speech, in Models for the perception of speech and visual form, Proceedings of a symposium, ed. by W. Wathen-Dunn, The M.I.T. Press, s. 111-125, 1967
- FISCHER-JØRGENSEN, E., Acoustic analysis of stop consonants, Miscellanea Phonetica, II, s. 42-59, 1954
- What can the new techniques of acoustic phonetics contribute to linguistics? Proceed. 8th Congr. ling., s. 433-478, 1958

- HADDING-KOCH, ABRAMSON, Duration versus spectrum in Swedish vowels: some perceptual experiments, *Studia Linguistica*, XVIII:II, s. 94-107, 1964
- HALLE, HUGHES, RADLEY, Acoustic properties of stop consonants, *Journ. Acoust. Soc. Am.*, 29:1, s. 107-116, 1957
- HARRIS, HOFFMAN, LIBERMAN, DELATTRE, COOPER, Effect of third-formant transitions on the perception of the voiced stop consonants, *Journ. Acoust. Soc. Am.*, 30:2, s. 122-126, 1958
- HEINZ, LINDBLOM, LINDQVIST, Patterns of residual masking for sounds with speech-like characteristics, *STL-QPSR*, 2-3, s. 66-71, 1967
- JAKOBSON, HALLE, Fundamentals of language, The Hague, 1956
- JOHANSSON, K., Några perceptoriska försök med svenska dentaler och supradentaler, lic.avh., 1967
- Transitions as place cues for voiced stop consonants: direction or extent?, *Studia Linguistica* XXIII:II (under tryckning), 1969
- LADEFOGED, P., Three areas of experimental phonetics, 1967
- LIBERMAN, DELATTRE, COOPER, GERSTMAN, The role of consonant-vowel transitions in the perception of the stop and nasal consonants, *Psychological Monographs* No. 379, Vol. 68, No. 8, 1954
- LISKER, MÁRTONY, LINDBLOM, ÖHMAN, F-pattern approximations of voiced stops and fricatives, *STL-QPSR*, 1, s. 20-22, 1960
- VIEREGGE, W.H., Die akustische Struktur der Plosivlaute (Ein Literatur-Überblick), *Forschungsbericht* 66/3, Institut für Phonetik und Kommunikationsforschung, Abteilung Kommunikationsforschung, der Universität Bonn
- ÖHMAN, S., Coarticulation in VCV utterances: spectrographic measurements, *Journ. Acoust. Soc. Am.*, 39:1, s. 151-168, 1966.