

Percepció de la parla

Ferran Pons

PID_00151860



Universitat Oberta
de Catalunya

www.uoc.edu

Índex

Objectius.....	5
1. Introducció: característiques generals i propietats del senyal de la parla.....	7
1.1. El fenomen de la coarticulació	8
1.2. El problema de la segmentació i la manca d'invariància	9
2. Descriptors: les unitats bàsiques de la parla.....	10
2.1. Trets articuladoris (fonètics)	10
2.2. Segments vocàlics i consonàntics	11
3. Processos perceptius.....	13
3.1. Percepció categòrica	13
3.2. Organització interna de les categories vocàliques: l'efecte imant	16
3.3. Reorganització de les categories consonàntiques	18
3.4. Les unitats de la percepció de la parla	18
4. La percepció de la parla des d'una perspectiva intermodal....	20
4.1. L'efecte McGurk: una demostració clàssica de la integració audiovisual de la parla	20
4.2. Els inicis de la integració audiovisual de la parla	21
5. Models teòrics.....	22
5.1. La teoria motriu	22
5.2. Teories auditives	23
Bibliografia.....	25

Objectius

Els objectius bàsics que ha d'assolir l'alumnat són els següents:

- 1.** Aprendre i comprendre els conceptes i termes bàsics corresponents als processos perceptius implicats en la percepció de la parla.
- 2.** Aprendre i comprendre l'especificitat del senyal de la parla.
- 3.** Saber identificar i ubicar els components i mecanismes dels processos perceptius des dels models teòrics.
- 4.** Familiaritzar-se amb la recerca i el mètode científic, a partir d'entendre els diferents experiments que s'expliquen sobre diferents processos i fenòmens en percepció de la parla.
- 5.** Raonar científicament, entendre el valor de les dades empíriques i la relació que tenen amb les explicacions teòriques.

1. Introducció: característiques generals i propietats del senyal de la parla

La percepció de la parla és un fenomen fascinant i molt complex al mateix temps. El procés de la percepció de la parla implica que a partir d'una sèrie de turbulències generades en expulsar l'aire dels pulmons a través de la laringe i el pas per les cordes vocals es produeixen una sèrie d'idees o conceptes en l'individu que ens escolta. En aquest mòdul intentarem donar una idea general de quins són els problemes bàsics i els processos bàsics que hi ha al darrere de la percepció de la parla.

Quan es pensa en el llenguatge pot ser que es pensi tant en el llenguatge oral com en el llenguatge escrit.

El llenguatge escrit és una invenció de l'home, mentre que el llenguatge oral (la parla) és una propietat innata de l'ésser humà.

La distinció entre **llenguatge oral** i **llenguatge escrit** no és crucial per a estudiar determinats aspectes del llenguatge, com pot ser el processament de la sintaxi o l'estudi del significat de les paraules. Tot i això, pels problemes que tractarem a continuació és fonamental distingir entre aquests dos tipus de llenguatges. La majoria de les nostres interaccions es produeixen mitjançant el llenguatge oral; l'aprenentatge de la parla durant els primers anys de vida succeeix sense tenir coneixement de l'escriptura.

Una de les característiques principals que diferencia el llenguatge escrit i el llenguatge oral és que en el primer es poden observar límits en les paraules (mitjançant els espais en blanc).

No és massa difícil per a qualsevol persona saber quantes paraules hi ha en una frase escrita, ja que es pot fer comptant el nombre d'unitats separades per espais; per exemple, en la frase de la figura 1, *El gos ha estat jugant amb el nen*, podem comptar fàcilment el nombre de paraules que la constitueixen: vuit paraules. En el senyal de parla, en canvi, no existeix cap límit clar entre paraules: no hi ha silencis o **espais en blanc** que ens permetin saber els límits de les paraules. La imatge de la figura 1 representa l'ona acústica corresponent a la frase *El gos ha estat jugant amb el nen*.



Figura 1
Representació de l'ona acústica corresponent a
El gos ha estat jugant amb el nen.

Com podem comprovar a primera vista no és tan simple i obvi saber quantes paraules formen aquesta ona, cosa que demostra que no existeix una correspondència entre els límits de les paraules i els fragments sense so en el senyal

de la parla. Tot i això, podem identificar **pauses** o "espais" en l'ona acústica de la figura 1; aquestes pauses es fan normalment per a respirar o per a planificar el contingut del que s'explica, però és important tenir en compte que aquests "espais" no delimiten cadascuna de les paraules, i així la parla és caracteritzada com un continu sonor.

A més a més d'aquesta diferència general entre llenguatge oral i escrit, hi ha altres diferències molt importants. Tinguem en compte que, si l'única diferència entre aquests dos llenguatges fos la descrita anteriorment, el llenguatge oral seria equivalent al llenguatge escrit però sense espais en blanc. Així, doncs, existeixen altres característiques importants que difereixen entre aquests dos tipus de llenguatges. Una d'aquestes característiques és que, a diferència de les lletres del llenguatge escrit, en el llenguatge oral els sons fonamentals que el formen (de moment considerem que són els **fonemes**) no estan separats de manera clara i nítida com succeeix amb les lletres, i fins i tot moltes vegades un mateix fonema no sona de la mateixa manera.

En resum, quan percebem la parla ens adonem que no existeix una frontera o un límit clar entre les paraules o fins i tot entre els sons que la conformen.

Exemple

Un exemple clar d'això el trobem en les dificultats que tenim de percebre correctament els sons o les paraules d'una llengua estrangera que desco-neixem.

1.1. El fenomen de la coarticulació

Com ja s'ha avançat, els sons que conformen la parla (fonemes) no es produeixen sempre de la mateixa manera. Observem la manera en què posem la llengua quan pronunciem les dues /n/ en les paraules *cantar* i *cangur*. La posició de la llengua és completament diferent en cada cas. Aquest fet passa perquè, en pronunciar la /n/, "anticipem" la posició del fonema que la segueix (/t/ i /g/ respectivament). Hem de ser conscients que aquest fenomen succeeix perquè mentre parlem els òrgans encarregats de produir la parla es mouen de manera contínua, i passen gradualment d'una posició necessària per a produir un determinat so a una altra posició, que és l'adequada per a produir el so que ve a continuació. Així, doncs, quan parlem coarticulem els diferents fonemes o sons de la parla. El fenomen de la **coarticulació** té una altra conseqüència en el senyal de la parla: existeix un encavalcament molt important dels sons de la parla. La continuïtat dels moviments del tracte vocal fa que en un mateix moment hi hagi informació corresponent a més d'una unitat lingüística.

Aquest fenomen ha estat descrit com a **linealitat**.

1.2. El problema de la segmentació i la manca d'invariància

Amb tota la informació proporcionada fins ara, podem ser conscients que la tasca de segmentar la parla no és una comesa senzilla: per una part, la dificultat es troba en el fet que no hi hagi separacions entre les unitats lingüístiques que conformen el missatge i, per l'altra, pot ser que un mateix so es pronunciï de manera diferent.

El fet que els sons no es produeixin sempre de la mateixa manera és el que es coneix com a **variació fonètica**. Els sons es pronuncien de manera diferent en funció de si es parla més ràpid o més a poc a poc, si el so és accentuat o no, entre altres factors.

També, des d'un punt de vista acústic, es pot observar que les realitzacions són diferents depenent de la veu: una veu d'un home, d'una dona o d'un nen dóna com a resultat diferències notables en la producció de cada so. Aquest fenomen es coneix com a **problema de la invariància**.

Així, doncs, com que la informació de cada segment fonètic és distribuïda en el senyal, encavalcant-se i fins i tot confonent-se a vegades entre aquests, Chomsky i Miller (1963) van plantejar que els fonemes havien de complir dues condicions formals importants –per a poder ser la base a partir de la qual es pogués construir una teoria fonològica–: la invariància i la linealitat.

Condicció de linealitat: per a cada fonema ha d'existir una porció de so en la producció i, de la mateixa manera que el fonema /b/ precedeix el fonema /a/ en la síl·laba /ba/, sense que existeixi encavalcament, en la banda sonora tampoc no s'han de produir encavalcaments entre la porció corresponent a /b/ i la corresponent a /a/.

Condicció d'invariància: per a un determinat fonema ha d'existir un conjunt d'atributs acústics específics i crítics, o trets associats en tots els contextos.

En resum, els problemes als quals s'enfronta l'organisme per a poder recuperar un missatge a partir d'un senyal acústic són principalment dos: falta d'invariància i de segmentació de l'estímul sonor: l'estímul no es presenta ni invariant ni apropiadament segmentat pel que fa al missatge lingüístic que es pretén recuperar. Tot i això, per a qualsevol parlant, les paraules, les síl·labes o inclús els fonemes es perceben clarament separats, tot i saber que físicament no són iguals.

2. Descriptors: les unitats bàsiques de la parla

La parla es pot definir a partir de les seves propietats acústiques (freqüència i amplitud) i temporals (duració). Les variacions en qualsevol d'aquestes dimensions comporten diferències en l'estímul sonor, que en alguns casos poden ser rellevants per a la percepció de la parla. La unitat mínima de so que serveix per a distingir significats ha estat definida com a fonema. Per a descobrir els fonemes d'una llengua, hem de buscar en les paraules els elements específics que en variar produeixen un canvi en el significat.

Per exemple, tenint en compte les paraules, *bala*, *sala*, *cala*, i *pala*, podem saber que els sons que representen les lletres inicials de cada paraula *b*, *s*, *c*, i *p* corresponen a fonemes del català.

2.1. Trets articuladoris (fonètics)

Podem considerar els fonemes com la unitat mínima de so que permet distingir significats, o també podem considerar els fonemes com una agrupació o combinació de trets subfonèmics, que són relativament independents entre si.

D'aquesta manera, per exemple, el fonema /p/ està compost pels trets consonant, sord, bilabial i oclusiu, mentre que el fonema /t/ està compost pels trets consonant, sord, dental i oclusiu.

El fet de considerar els **trets subfonèmics** com a unitats independents de la parla té una sèrie de conseqüències.

- La primera és que algunes de les distincions entre els sons haurien de ser les mateixes.

És a dir, si /b/ i /p/ difereixen només en sonoritat, de la mateixa manera que passa amb /t/ i /d/, les diferències existents entre /b/ i /p/ haurien de ser iguals a les existents entre /t/ i /d/.

- La segona és que, si els trets subfonèmics funcionen com a unitat, s'establirien diferències entre fonemes.

Per exemple, si /b/ i /p/ difereixen en sonoritat, i /b/ i /d/ en el punt d'articulació, mentre que /b/ i /t/ difereixen en sonoritat i punt d'articulació, llavors /b/ i /t/ haurien de ser més diferents que /b/ i /p/ o que /b/ i /d/, ja que els separen dos trets i no solament un.

- Finalment, se suposa l'existència d'una jerarquia de trets, que implicaria que algunes distincions són més importants que d'altres.

Així, doncs, tot i que /b/ i /m/, i /b/ i /p/ es distingeixen per un únic tret (forma d'articulació en la primera, i sonoritat en la segona), ja que la forma d'articulació és més important, /b/ i /m/ haurien de ser discriminables més fàcilment que /b/ i /p/.

Jerarquia de trets

Hi ha diversos tipus d'estudis que han donat suport a una jerarquia de trets. Un d'aquests tipus d'estudis són els estudis de semblança fonètica. Els estudis de semblança consisteixen a fer que el participant valori la semblança de diferents fonemes a un de determinat. Per exemple, es demana als subjectes que ordenin segons la semblança al fonema /p/ els fonemes /m/, /s/, /d/, /a/, /b/ i /r/. Els resultats que s'obtenen indiquen que l'ordre que s'estableix de semblança amb el fonema escollit és determinat pel nombre de trets que tenen en comú aquests fonemes amb el fonema a què es compara. La jerarquia de trets queda descrita de la manera següent: la distinció més important és la de vocal/consonant, a continuació hi trobem la forma d'articulació, després la sonoritat i finalment el punt d'articulació. És important remarcar que en aquest tipus d'estudis no s'observa la premissa d'independència de trets subfonètics, ja que s'observa que els participants perceben com a més importants les diferències en el punt d'articulació quan es tracta d'oclusives sordes que no pas quan es tracta d'oclusives sonores.

2.2. Segments vocàlics i consonàntics

La distinció més important entre els sons que formen la parla és la **distinció consonant-vocal**.

En general les consonants es caracteritzen perquè en l'expulsió de l'aire el conducte bucal imposa una sèrie de restriccions, mentre que les vocals es caracteritzen per una sortida relativament lliure de l'aire.

Les **consonants** es poden descriure utilitzant tres dimensions articulatòries:

- La **forma o mode d'articulació** es refereix a la naturalesa de la restricció, a la quantitat d'aire que es deixa passar i a la manera com surt a l'exterior. Segons aquesta dimensió, les consonants es classifiquen en oclusives [b, d, g, p, t, k], africades [tʃ], fricatives [f, ʃ, s, x, z], nasals [m, n] i líquides [l, r, rr].
- La **sonoritat** es refereix al fet de si les cordes vocals vibren o no mentre es produeix el fonema en qüestió. El so /z/ (com en la paraula *zinc*) és un so sonor, mentre que el so /s/ és un so sord (com en la paraula *cinc*). Alguns altres exemples són les oclusives sonores (/b/, /d/, /g/) i les oclusives sordes (/p/, /t/, /k/).
- El **punt d'articulació** es refereix al lloc de la cavitat bucal en què es produeix la restricció en la sortida de l'aire. Pot ser bilabial (es produeix als llavis), labiodental, dental, interdental, alveolar, palatal o velar.

Les **vocals** són totes sonores; l'aire flueix sense restriccions. La forma d'articulació no és rellevant, ja que no es produeix cap tancament que dificulti la sortida de l'aire. Per a aquest tipus de sons, la **localització de la base de la llengua** (la part més posterior, oposada a la punta de la llengua) és crucial:

- **El desplaçament horitzontal de la llengua:** si avança (vocals anteriors, com la [i]) o retrocedeix (vocals posteriors, com la [u]).
- **La posició vertical de la llengua:** si s'eleva (com les vocals [i] i [u]) o descendeix (com la vocal [a]). Un últim tret és l'arrodoniment dels llavis en el moment de l'emissió ([u] i [o] són vocals arrodonides, mentre que [i] i [e] no ho són).

Cada vocal es caracteritza, almenys, per tres **freqüències formants** (F1, F2 i F3, de més greu a més aguda). És important tenir en compte que no totes les realitzacions de les diferents vocals tenen els mateixos valors absoluts, ja que aquests canvien segons la veu (per exemple, home enfront de dona). Així, doncs, el que determina cada vocal és el patró espectral, o el que és el mateix, la relació relativa de les freqüències formants (f1, f2 i f3) que la conformen.

3. Processos perceptius

3.1. Percepció categòrica

Sovint, en percebre objectes o successos del nostre entorn ho fem mitjançant categories. Podem identificar un objecte o un altre basant-nos en la categoria a la qual pertany, però normalment les fronteres entre categories no són abruptes. Si pensem en els colors, som capaços de denominar o etiquetar els colors (vermell, verd, blau, etc.), però aquestes categories són contínues; podem percebre canvis quantitius en un color i mitjançant un continu passar del color vermell al taronja sense haver pogut ubicar una frontera categòrica inequívoca. En la percepció de la parla aquest fet canvia, ja que la percepció dels fonemes es regeix per un fenomen denominat **percepció categòrica**.

El fenomen de la percepció categòrica es refereix a la impossibilitat de discriminar sons de la parla millor que identificar-los: resulta més difícil, o inclús impossible, diferenciar sons que s'han identificat com a pertanyents a una mateixa categoria, tot i que siguin acústicament diferents. En canvi, la identificació d'aquests sons com a pertanyents a una determinada categoria és una tasca molt senzilla.

Exemple

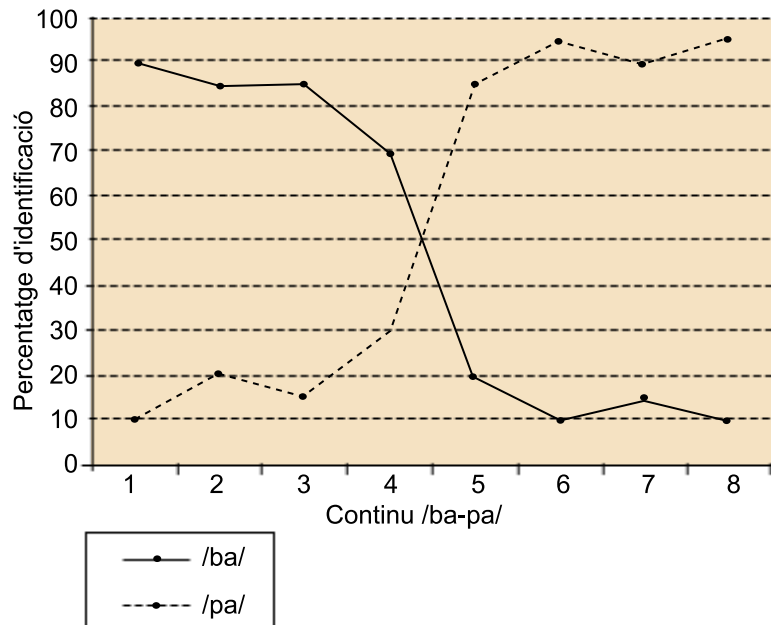
Per exemple, diferenciar entre dos sons de la consonant /p/ ens resulta molt difícil, mentre que identificar aquests sons com a /p/s és molt fàcil.

La percepció categòrica ha estat un fenomen àmpliament estudiat en el camp de la percepció de la parla, ja que reflecteix la naturalesa dels sons de la parla i la manera específica en què el nostre sistema perceptiu els tracta.

L'habilitat per a categoritzar els sons de la parla és un requisit essencial perquè l'oient compregui el discurs. Per a comprendre el llenguatge oral és necessari imposar una percepció categòrica al senyal i no simplement una determinació relativa de diverses característiques físiques d'aquest senyal. La noció de percepció categòrica va sorgir en experiments dissenyats a determinar com classificaven les consonants oclusives els parlants (Liberman, 1957). A partir d'aquests resultats inicials es va començar a explorar més a fons aquest fenomen.

Per a estudiar experimentalment el fenomen de la percepció categòrica es crea un continu de sons utilitzant parla sintetitzada, variant gradualment un determinat paràmetre. Per exemple, en un continu /ba-pa/ es manipula el temps d'inici de la vocalització (TIV). Aquest paràmetre, definit com la relació entre el moment de l'expulsió de l'aire i la vibració de les cordes vocals, pot oscil·lar des d'un valor negatiu (-150 ms aproximadament) fins a un valor positiu (+150 ms aproximadament). Tot i que el TIV varia de manera contínua, la percepció ho fa de manera abrupta. Existeix un límit invariable per a cada llengua a partir del qual es modifica la percepció d'un fonema en concret (deixa de ser /ba/ i es percep com a /pa/), i molt rarament es percep un so intermediari entre dos fonemes. Així, si es demana a un oient que categoritzi els sons d'un continu /ba-pa/, classifica els sons amb un TIV negatiu o positiu curt (< +10 ms) com a /ba/, mentre que els sons amb un TIV positiu llarg (> +10 ms) els classifica com a /pa/. El punt del continu on canvia la identificació d'un fonema per un altre es denomina *límit o frontera categorial*.

Figura 2

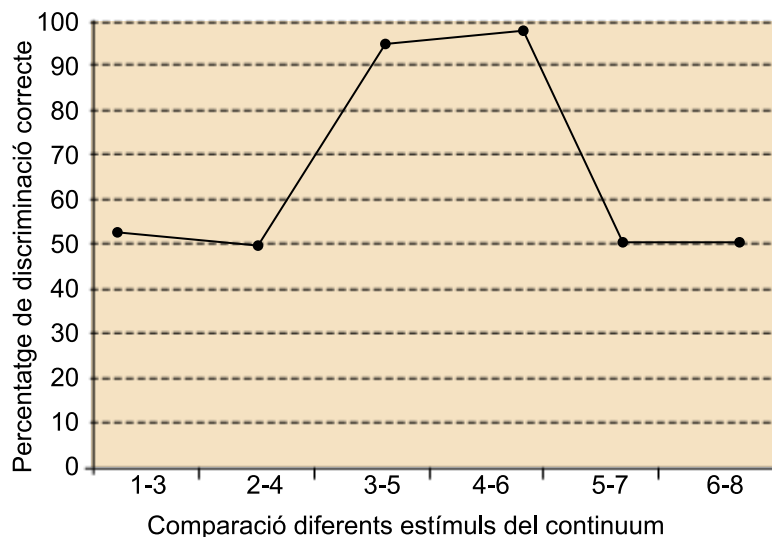


Resultats (hipotètics) d'una prova d'identificació que mostren la classificació dels estímuls en dues categories. La frontera categorial se situaria entre l'estímul 4 i 5.

A més a més de la tasca d'identificació (categorització) dels estímuls, existeix una segona tasca en percepció categòrica que consisteix en una **tasca de discriminació** (tasca ABX).

La tasca ABX consisteix a presentar a l'oient tres estímuls: els dos primers diferents i un tercer que pot ser igual al primer o igual al segon. Els participants han de dir quin dels dos primers estímuls és igual al tercer. La tasca és molt senzilla quan els dos primers estímuls pertanyen a dues categories fonètiques diferents, però quan els dos primers estímuls pertanyen a una mateixa categoria fonètica la tasca es complica i s'obtenen majoritàriament respostes a l'atzar (vegeu la figura 3).

Figura 3



Resultats (hipotètics) d'una prova de discriminació fonètica (ABX). La frontera se situaria en els estímuls 4-5.

Hem observat gràcies a tota l'evidència experimental dels últims 30 anys que els fonemes, especialment les consonants, estan organitzats categòricament i que es regeixen pel fenomen de la percepció categòrica. Així, doncs, sabent com estan organitzats els fonemes, encara podem preguntar-nos: com es formen aquestes categories fonètiques? **El fenomen de la percepció categòrica, és present des dels primers mesos de vida o s'adquireix més tard mitjançant l'experiència amb la llengua materna?**

Per a respondre a aquesta pregunta Eimas i col·laboradors (1971) van estudiar el fenomen de la percepció categòrica en nadons d'edats compreses entre 1 i 4 mesos. En l'estudi van observar la discriminació d'una sèrie d'estímuls del contínuum /ba-/pa/ que s'obtenien modificant-ne el temps d'inici de la vocalització (TIV). Com ja s'ha explicat, els adults no són capaços de percebre les diferències en els TIV entre dos estímuls, si no és que les dues síl·labes que es comparen pertanyen a dues categories fonètiques diferents, és a dir, que una sigui la síl·laba /ba/ i l'altra, la síl·laba /pa/. Així, doncs, si s'observa que els nadons tenen el mateix patró de discriminació, s'obtindria evidència que perceben el contínuum en dues categories de la mateixa manera que ho fan els adults.

Per a explorar la resposta dels nadons enfront d'aquests estímuls fonètics es va utilitzar la **tècnica de succió d'alta amplitud** o no nutritiva, basada en la preferència que mostren els nadons pels estímuls nous.

El procediment consisteix a posar el nadó en una hamaca petita situada davant d'uns altaveus i donar-li un xumet per succionar. Aquest xumet és connectat a un ordinador i es mesura el nombre de succions per minut i la força o amplitud d'aquestes succions. S'estableix un període anomenat *d'habitució* en el qual el nadó escolta repetidament un mateix estímul fins que la taxa de succió disminueix o s'estabilitza, i mostra així un "desinterès" per l'estímul. Seguidament es presenta un estímul nou i s'observa la taxa de succió. Si el nadó percep l'estímul com a diferent, la taxa de succió augmenta.

Eimas i col·laboradors (1971) van presentar als nadons una sèrie d'estímuls del continuu /ba/ i /pa/ que diferien en el TIV en 20 ms. Es van posar a prova dos tipus de parells d'estímuls: el parell d'estímuls "iguals", que consistien en estímuls que diferien en el TIV de 20 ms però que pertanyien a la mateixa categoria /ba/ o /pa/ (segons resultats obtinguts amb adults), i els parells "diferents" (D), que consistien en dos estímuls que diferien en el TIV de 20 ms i que a més a més pertanyien a dues categories diferents. Els resultats van indicar que els nadons augmentaven la taxa de succió únicament quan els estímuls que es comparaven pertanyien a categories fonètiques diferents.

L'efecte de la percepció categòrica dels fonemes ja s'observa al començament de la vida de l'ésser humà.

3.2. Organització interna de les categories vocàliques: l'efecte imant

Una de les teories més influents sobre els mecanismes implicats en la reorganització perceptiva dels sons de la parla va ser desenvolupada per **Patricia Kuhl (2000)**. La **teoria de l'imant de la llengua nativa (TILN)** aborda, entre altres aspectes, el període primerenc de la percepció de la parla (abans que els nadons adquireixin el significat de les paraules i la fonologia contrastiva), i també els canvis en les habilitats en la percepció de la parla produïdes per l'experiència amb la llengua.

La TILN és basada en la hipòtesis del prototip, que prediu que tots els membres d'una categoria no són percebuts com a equivalents, ja que alguns membres es perceben com a millors exemplars de la categoria que d'altres.

Si les categories vocàliques són estructurades inicialment, la tipicitat de l'exemplar usat com a referent per a la categoria vocàlica afecta la percepció dels altres membres de la categoria. Específicament, si el prototip és més representatiu d'una categoria, llavors els altres membres de la mateixa categoria es perceben com a més similars a aquests en comparació d'un estímul no prototípic. Així, es va observar que els prototips fonètics actuen de manera especial en la percepció de la parla, ja que actuen com a "imants perceptius" per als altres sons de la seva mateixa categoria fonètica. Quan percebem el prototip d'una categoria fonètica i el comparem amb sons que el rodegen en l'espai acústic, ens resulta difícil, ja que aquest prototip actua com un *imant*, de manera que fa més difícil la discriminació entre aquest i altres exemplars propers. En canvi, quan percebem un so no prototípic i l'hem de comparar amb algun dels sons que el rodegen, la discriminació és molt més fàcil.

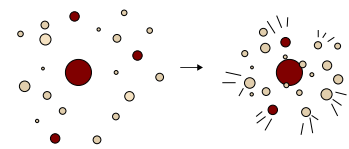
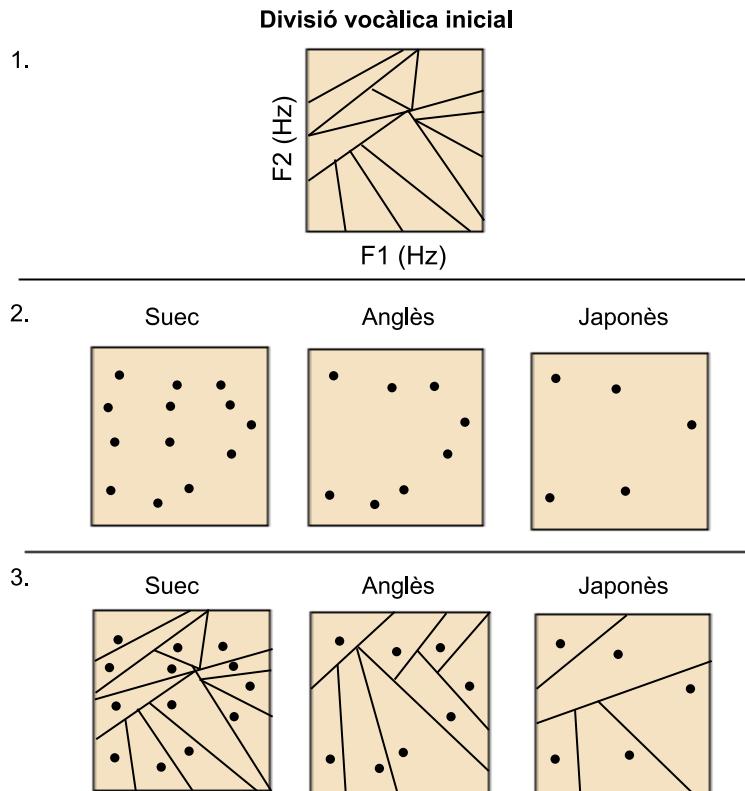


Figura 4
L'efecte imant perceptiu. Els exemplars més propers al prototip es confonen amb aquest prototip, són percebuts com el mateix.

Així, doncs, un dels fenòmens que sustenta la TILN és l'efecte imant perceptiu. Se suggereix que inicialment el sistema de representació auditiva està dividit de tal manera que els nadons són capaços de distingir qualsevol diferència fonètica rellevant de qualsevol llengua. Aquesta habilitat queda il·lustrada en un hipotètic espai vocàlic basat en coordenades dels dos primers formants que defineixen les vocals (F1 i F2). Les divisions inicials d'aquest espai vocàlic separen perceptivament un nombre ampli de vocals (figura següent, part 1).



Aquesta capacitat per a percebre els sons vocàlics dividits en múltiples categories separades per límits naturals és present des del naixement. Tot i això, durant el primer any de vida, i a causa de l'exposició a la parla, l'espai fisicoacústic de la llengua es modifica. Es formen els prototips de les categories fonètiques i aquests prototips van contraient l'espai perceptiu, i atreuen els membres de la categoria cap a ells. Aquest fenomen ha estat definit com a **efecte imant perceptiu**. Conseqüentment, existeix una pèrdua de discriminabilitat prop d'aquests valors prototips de la pròpia llengua. Així, per exemple, l'espai vocàlic als 6 mesos d'edat queda modificat de manera sorprenent. En tres ambients lingüístics diferents, com anglès, suec i japonès, podem observar una distribució vocàlica totalment diferent (figura anterior, part 2). Els nadons ja disposen de representacions mentals dels sons de la parla i han format els mapes perceptius específics de la seva llengua materna (figura anterior, part 3) (Kuhl i col·laboradors, 1992).

3.3. Reorganització de les categories consonàntiques

L'observació feta per Kuhl i col·laboradors sobre l'organització de les vocals de la llengua materna durant el primer any de vida no és exclusiva dels sons vocàlics, ja que també s'ha observat un patró *similar* en les consonants. Werker i Tees (1984) van explorar la percepció dels sons consonàntics durant el primer any de vida. Van realitzar un estudi per a explorar el curs temporal en la pèrdua en l'habilitat de discriminar sons consonàntics que no pertanyen a la llengua materna. Per això van explorar les habilitats discriminatives de nadons nascuts en entorns de parla anglesa, d'edats compreses entre 6-8 mesos, 8-10 mesos i 10-12 mesos, utilitzant contrastos consonàntics de la llengua hindi (/Ta-ta/) i de la llengua salish (/k'i-q'i/). Utilitzant el **procediment de gir de cap** (en anglès *conditioned head-turn paradigm*) van analitzar la capacitat de discriminació dels nadons enfront d'aquests parells de contrastos consonàntics de llengües no maternes.

Els resultats van mostrar que els nadons més petits (de 6 a 8 mesos) podien discriminar ambdós contrastos fonètics no materns amb facilitat, mentre que els nadons més grans (de 10 a 12 mesos) pràcticament no presentaven aquesta habilitat discriminativa, i tan sols uns quants nadons eren capaços de discriminar aquests contrastos no materns.

Els resultats d'aquest estudi donen suport a la hipòtesi que els nadons inicialment poden discriminar qualsevol contrast fonètic d'una llengua natural, i que existeix un declivi en aquesta habilitat com a conseqüència de l'exposició a la llengua materna.

El declivi per a discriminar sons consonàntics que no pertanyen a la llengua materna sembla que és evident cap al final del primer any de vida.

3.4. Les unitats de la percepció de la parla

Fins a aquest moment la unitat que hem tractat ha estat el fonema. Com hem dit, no és erroni afirmar que el fonema és la unitat a partir de la qual es forma la parla. Tot i això, els segments que formen la parla no necessàriament han de ser aquestes unitats fonètiques. Les unitats que extraïem en percebre la parla, són els trets acusticofonètics explicats anteriorment? O potser són els fonemes? Una altra possibilitat que encara no hem previst és la unitat de la **síl·laba**. En aquest apartat examinarem les unitats a les quals accedim en percebre la parla.

La importància relativa dels fonemes i les síl·labes a l'hora d'accedir a la parla ha estat un tema de debat en les últimes dècades. A primera vista, sembla obvi pressuposar que l'ordre lògic és de primer identificar els fonemes i que poste-

riorment aquests fonemes s'agrupin en síl·labes. Tot i això, amb tot el que s'ha explicat fins ara sobre els fonemes (per exemple, els fenòmens d'invariància o linealitat), ens podríem plantejar que el procés pot ser inclús a l'inrevés.

Un primer estudi de Liberman i col·laboradors (1974) va explorar la capacitat de nens de diferents edats de manipular fonemes i síl·labes. Van observar que no tenien cap problema a segmentar les paraules per síl·labes, però que la segmentació per fonemes es produïa a partir dels 6 anys. Per comprovar si era degut al paper de l'aprenentatge lector, Morais i col·laboradors (1979) van fer el mateix estudi amb adults analfabets i alfabetitzats. Els seus resultats van mostrar que els adults analfabets podien segmentar en síl·labes però no en fonemes. Així, doncs, sembla que l'adquisició del concepte de fonema seria deguda a l'adquisició d'un sistema de lectura determinat i no al desenvolupament normal del sistema lingüístic.

En una altra sèrie d'experiments, utilitzant la tècnica de detecció de sons, Mehler i col·laboradors (1981) van observar que els seus participants francesos eren més ràpids a detectar segments quan aquests segments coincidien amb la síl·laba que quan no hi coincidien (per exemple, van mostrar temps de reacció més baixos, és a dir, van ser més ràpids, a detectar la síl·laba /pa/ en paraules com *palace* ('palau') que en paraules com *palmier* ('palmera'), mentre que el temps de detecció va ser inferior a detectar /pal/ a *palmier* que a *palace*). Els resultats es van interpretar com que els participants segmentaven el senyal de parla en unitats que corresponien a la síl·laba. Tot i això, aquest resultat no es va poder replicar amb participants anglesos (Cutler i col·laboradors, 1983). Aquests últims, mostraven temps equivalents a detectar el segment, independentment que aquest segment pertanyés a una síl·laba o no. Es planteja que els subjectes utilitzen una estratègia de sil·labificació únicament quan la seva pròpia llengua és fàcilment segmentable en síl·labes (aquest no és el cas de l'anglès).

4. La percepció de la parla des d'una perspectiva intermodal

Fins ara hem descrit i explicat la percepció de la parla des d'una vessant acústica. És evident, però, que si ens hi mirem bé ens adonarem que la parla no és únicament i exclusivament acústica, sinó que també hi ha altres modalitats que intervenen, com la **modalitat visual**. És per això que la percepció de la parla s'ha d'explorar també des d'una perspectiva intermodal. De fet, ha quedat demostrat que els processos d'integració multisensorial de la parla tenen un paper molt important en la percepció de la parla cara a cara. Evidentment, com ja hem comentat, el senyal auditiu per si mateix és suficientment precís i detallat per a una comprensió molt bona de la parla (com queda palès quan utilitzem el telèfon o escoltem la ràdio). Sembla que per aquesta raó les propietats acústiques de la parla han constituït i constitueixen el nucli més important de la recerca en la percepció de la parla. Tot i això, el fet de poder veure els **gestos visuals** o **articulatoris** de la parla fa que hi hagi un benefici a l'hora de percebre-la i interpretar-la. Diversos estudis ho demostren.

Per exemple, Sumbly i Pollack (1954) van demostrar que, quan es perd una part de la intel·ligibilitat d'un missatge verbal acústic afegint-hi soroll, la presentació concurrent de les corresponents pistes visuals de la parla millora la comprensió de la parla espectacularment, a un nivell equivalent a un increment acústic de 15-20 decibels. La informació visual de la parla no solament dóna pistes redundants per a enfortir l'estímul auditiu, sinó que també ajuda a no fer tan ambigus alguns sons de la parla que difereixen en el mode d'articulació però que sonen molt similars. De la mateixa manera Reisberg i col·laboradors (1987) van demostrar que, en escoltar un missatge perfectament audible d'una persona amb un accent estranger molt fort o un missatge amb un contingut semàntic complicat, la presentació de la informació visual ajudava a comprendre'l. Fins i tot, s'ha observat que a l'hora de percebre els fonemes d'altres llengües que s'estan aprenent, la informació visual de la parla en pot facilitar la percepció (Navarra i Soto-Faraco, 2007).

4.1. L'efecte McGurk: una demostració clàssica de la integració audiovisual de la parla

La teoria motora (explicada al final d'aquest mòdul) postula que existeix un vincle entre la percepció i la producció del llenguatge: la informació sobre els gestos articulatoris és la base de la percepció de la parla. Així, doncs, la informació visual sobre l'articulació s'hauria d'integrar en la informació auditiva. L'efecte McGurk (McGurk i MacDonald, 1976) demostra aquest fenomen.

Per a observar el fenomen es demana als participants que transcriguin els sons que perceben. Existeixen tres condicions experimentals: la visual, en la qual els participants veuen una persona que produeix les síl·labes sense escoltar res; l'auditiva, en la qual els participants escolten síl·labes, però no veuen la cara de la persona, i finalment l'audiovisual, en la qual escolten les síl·labes i veuen la persona pronunciar-les. En l'última condició, en alguns casos la informació visual no es correspon amb la informació auditiva (per exemple, es presenta auditivament /ba/ però la cara gesticula la síl·laba /ga/). Com que existeix un vincle entre la producció i la percepció de la parla, succeeix que la nostra percepció es basa en la mescla de la informació visual i de l'auditiva. Així, doncs, quan la síl·laba que es presenta auditivament no concorda amb la que veiem, el nostre sistema les uneix, i crea així la percepció d'una síl·laba nova; per tant, si un individu escolta la síl·laba /ba/, i al mateix temps veu la persona pronunciant la síl·laba /ga/, és molt probable que la percepció final sigui la de la síl·laba /da/.

4.2. Els inicis de la integració audiovisual de la parla

Hem observat que en la percepció de la parla integrem la informació visual i auditiva. Existeixen diversos resultats amb participants adults que ho demostren. Tot i això, encara ens queda per resoldre la qüestió de si els nadons són capaços d'integrar la informació auditiva i visual de la parla des d'etapes primerenques o si al contrari potser aquesta correspondència entre el gest visual i el so s'aprèn molt més tard. Tot i que no existeix una evidència directa que els nadons de pocs mesos puguin integrar la informació audiovisual de la parla, hi ha alguns estudis que mostren que tenen nocions sobre la correspondència entre els moviments articuladoris que es produeixen a la cara i la informació sonora o el senyal auditiu que hi correspon. Per exemple, Kuhl i Meltzoff (1984) van explorar la capacitat de nadons de 4 a 5 mesos per a veure la correspondència auditiva i visual de dues vocals de l'anglès.

El procediment implicava dues fases: en la primera fase, no es presentava cap so i els nadons veien de manera repetida dues cares, una al costat de l'altra, i cada cara articulava una vocal diferent; en la segona fase, es presentaven les mateixes cares articulant les vocals i al mateix temps una de les dues vocals era presentada acústicament. Els resultats van mostrar que els nadons atendien més estona a la cara que corresponia al so de la vocal presentada acústicament. Aquests resultats indiquen que els nadons tenen un coneixement implícit de la relació entre la informació visual i auditiva de la parla.

Patterson i Werker (2003) van observar que inclús nadons de dos mesos percebien la correspondència entre els sons vocàlics i el gest articuladori corresponent, cosa que demostra que el fenomen de la integració audiovisual dels sons de la parla en nadons és un fenomen sòlid.

5. Models teòrics

5.1. La teoria motriu

La teoria motriu de la percepció de la parla va ser postulada a finals dels anys seixanta per Liberman i els seus col·laboradors als laboratoris Haskins dels Estats Units. Van proposar que la percepció de la parla no es feia directament a partir del senyal acústic, sinó a partir de la producció (articulació).

La percepció de la parla és possible ja que tenim un mecanisme especial, totalment diferent del mecanisme general encarregat de la percepció auditiva.

Hi ha tres premisses o supòsits bàsics que sustenten aquesta teoria:

- La percepció de la parla és **basada en la producció o articulació**. El fet de ser capaços de produir la parla indica que coneixem de manera implícita com es produeixen els sons i com, per exemple, la coarticulació complica la correspondència entre el que un parlant vol dir i la forma final que adquireix. Així, doncs, hi ha un nexa inseparable entre el sistema responsable de percebre la parla i el de produir-lo.
- Se suposa que la percepció de la parla és **específica de l'espècie humana**. Adonem-nos que aquest segon supòsit deriva directament del primer, ja que l'ésser humà és l'únic ésser que parla i per tant l'únic que té coneixement sobre la manera com es parla. Per tant, si la percepció de la parla es basa en el coneixement de l'articulació dels fonemes, únicament els humans poden percebre la parla de manera estructurada.
- La **percepció de la parla és innata**. El sistema especialitzat de la percepció i la producció de la parla forma part del nostre patrimoni genètic. A escala anatòmica es porta a terme per àrees específiques del cervell (àrees de l'hemisferi esquerre). Així, doncs, se suposa que aquest coneixement de la parla ja és operatiu durant la primera infància.

S'han observat diversos fenòmens en la percepció de la parla a favor de la teoria motora, entre els quals, els més importants, i que ja s'han explicat anteriorment, hi ha l'efecte McGurk i la percepció categòrica de la parla.

5.2. Teories auditives

A diferència de la teoria motriu, les teories auditives es basen en el paper crucial dels mecanismes perceptius, i resten importància, doncs, a la producció o articulació de la parla.

El **model de percepció de lògica difusa** (*fuzzy logical model of perception*), de Massaro i col·laboradors, es basa en l'extracció d'indicis acústics, i remarca la importància de les diferents fonts d'informació, concretament de la relació entre la visió i la descodificació auditiva. La base metodològica d'aquest model se centra en la percepció bimodal (com en l'efecte McGurk, explicat prèviament).

En aquest model es proposa que l'individu recorda els prototipus dels fonemes (consistents en els trets acusticofonètics que caracteritzarien cada fonema) i, en percebre la parla, fa una estimació o comparació del so que escolta amb els prototipus emmagatzemats a la memòria.

Per a determinar el grau d'adequació del que percep amb el prototip, l'individu utilitza la combinació de les diferents entrades (auditiva i visual). Si una de les fonts d'informació és ambigua, l'altra pren un paper predominant sobre la decisió de la naturalesa de l'estímul.

Per altra banda, el **model lexical access from spectra** (LAFS), de Klatt, proposa que l'individu no segmenta el senyal de la parla en fonemes, sinó que a partir del flux sonor genera directament un "candidat" a paraula.

L'individu té emmagatzemat a la memòria *difonemes* –combinacions de dos sons (seqüències de vocal-consonant o consonant-vocal)–, que es comparen directament amb l'estímul lingüístic que arriba en forma de senyal sonor.

Bibliografia

Bibliografia bàsica

De Vega, M. i Cuetos, F. (1999). *Psicolingüística del Español*. Madrid: Trotta.

Coren, S., Ward, L. M., i Enns, J. T. (2001). *Sensación y Percepción* (5a. ed.). México, D.F.: Mc Graw Hill (pp. 369-392).

Pisoni, D. B. i Remez, R. E. (2005). *The handbook of speech perception*. Oxford, UK: Blackwell Publishing Ltd.

Goldinger, S. D., Pisoni, D. B., i Luce, P. (1996). Speech perception and spoken word recognition: Research and theory. A N. Lass (Ed.), *Principles of Experimental Phonetics* (pp. 277-327). Mosby St. Louis.

Referències bibliogràfiques

Cutler, A., Mehler, J., Norris, D., i Segui, J. (1983). A language-specific comprehension strategy. *Nature*, 304, 159-160.

Eimas, P. D., Siqueland, E. D., Jusczyk, P. W., i Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.

Chomsky, N. i Miller, G. A. (1963). Introduction to the formal analysis of natural languages. A R. D. Luce, R. R. Bush i E. Galanter (Eds.), *Handbook of Mathematical Psychology*, 2, 269-321. New York: Wiley.

Kuhl, P. K. (2000). Language, mind, and brain: Experience alters perception. A M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2a. ed., pp. 99-115). Cambridge, MA: MIT Press.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., i Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants six months of age. *Science*, 255, 606-608.

Liberman, A. M. (1957). Some results of research on speech perception. *Journal of the Acoustical Society of America*, 29, 117-123.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., i Carter, B. (1974). Reading and the awareness of linguistic segments. *Journal of Experimental Child Psychology*, 18, 201-212.

McGurk, H. i MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

Mehler, J., Dommergues, J. Y., Frauenfelder, U., i Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior*, 20, 298-305.

Kuhl, P. K. i Meltzoff, A. N. (1984). The intermodal representation of speech in infants. *Infant Behavior and Development*, 7, 361-381.

Morais, J., Bertelson, P., Cary, L., i Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, 7, 323-331.

Patterson, M. L. i Werker, J. F. (2003). Two-month old infants match phonetic information in lips and voice. *Developmental Science*, 6, 193-198.

Reisberg, D., McLean, J., i Goldfield, A. (1987). Easy to hear but hard to understand: a lip-reading advantage with intact auditory stimuli. A B. Dodd i R. Campbell (Eds.), *Hearing by Eye: the Psychology of Lip-reading Hillsdale*. New Jersey: Lawrence Erlbaum Associates.

Sumby, W. H. i Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26, 212-215.

Werker, J. F. i Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.

