

### La lettura degli spettrogrammi

Come discusso precedentemente, mentre il concetto di formante è ben definito in alcuni casi ideali, manca una precisa definizione operativa che permetta un'identificazione certa dei valori caratteristici assunti in diverse condizioni reali.

Si entra qui nei problemi dell'interpretazione dello spettrogramma, problemi che sono complessi per il fatto – messo ben in evidenza sin da Fant (1960) – che una stessa variante articolatoria, da un individuo all'altro e da un'esecuzione all'altra dello stesso parlante, può determinare manifestazioni spettrografiche diverse e che, reciprocamente, a un'indicazione spettrografica simile possono corrispondere strategie articolatorie diverse.

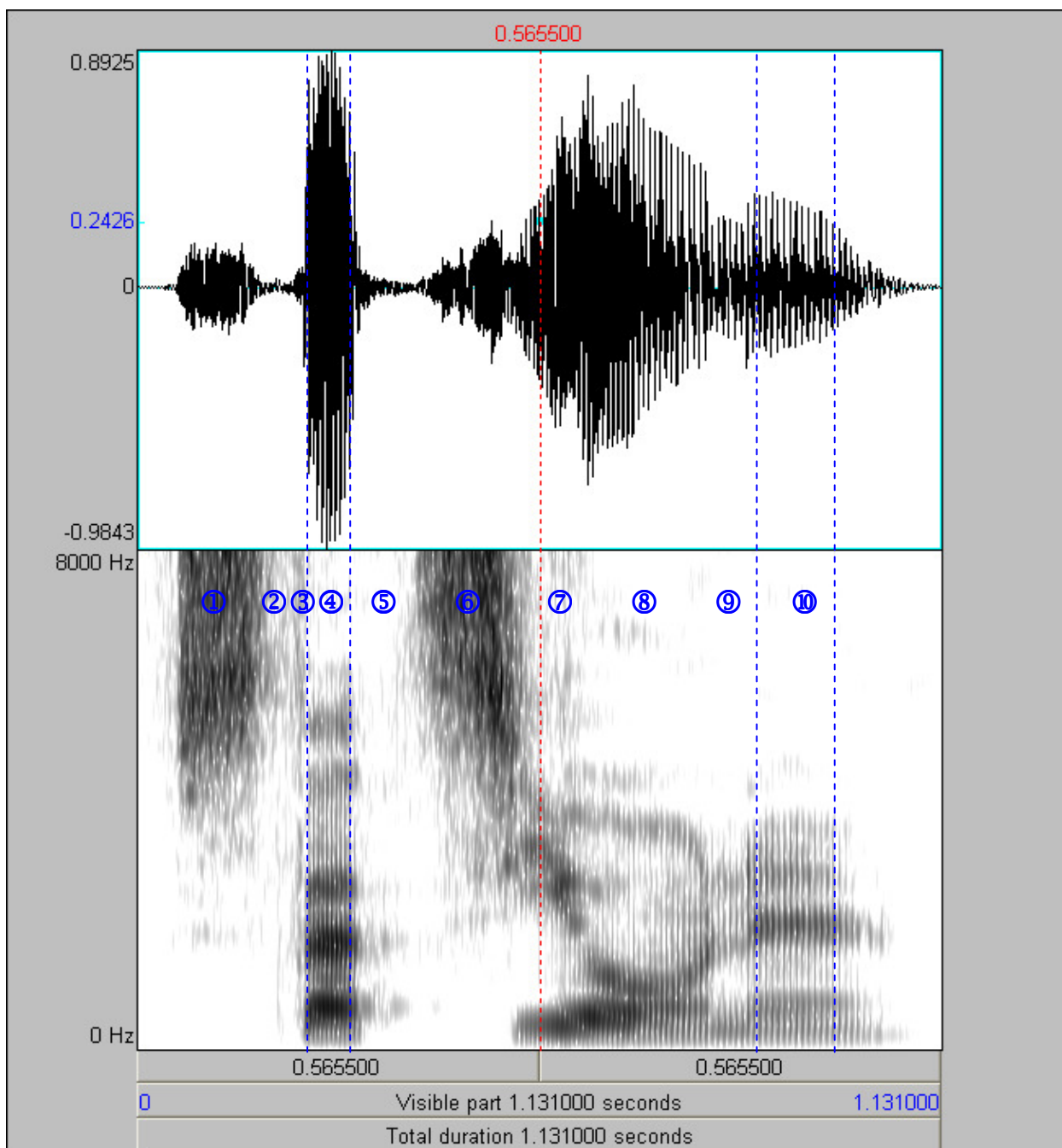


Figura C. Oscillogramma (in alto) e spettrogramma (in basso) di una realizzazione della parola *stazione* pronunciata da un parlante italiano (dicatore professionale). Si possono distinguere 10 fasi temporali (indicate dalle cifre cerchiare) associate a diversi momenti articolatori dei vari segmenti (v. testo).

L'analisi delle fasi in cui si presentano i diversi suoni di una parola inizia con la segmentazione del tracciato, cioè con l'individuazione delle caratteristiche delimitative dei segmenti e con la classificazione degli eventi presenti in ciascun segmento<sup>5</sup>.

Per riuscire a trarre dagli spettrogrammi il maggior numero d'informazioni relative al messaggio sottoposto ad analisi (quando è noto il codice usato dal locutore che lo ha prodotto) sono tuttavia necessarie alcune indicazioni di base. Ci proponiamo di discuterne qui alcune a partire dall'osservazione di un solo esempio relativo a una realizzazione della parola 'stazione' da parte di un parlante italofono senza particolari marche geo- o socio-fonetiche (v. Fig. C)<sup>6</sup>.

Le principali osservazioni che possiamo fare su questi grafici sono relativi alla particolare successione di segmenti che riusciamo a individuare nel corso dello sviluppo temporale della produzione (lungo l'asse delle ascisse), nonostante lo sviluppo continuo e gl'importanti fenomeni di coarticolazione.

Osserviamo sin dall'inizio, grazie alla bassa densità di righe verticali delle regioni di sonorità vocale, che si tratta di una voce maschile<sup>7</sup>.

Al di là di questo dato, facendo finta di non conoscere il contenuto sonoro (e d'ignorare quindi la rappresentazione fonetica simbolica) di questa produzione, si possono distinguere i segmenti acustici che si candidano a rappresentare i suoni dell'ideale catena sonora che la contraddistingue.

Notiamo da sinistra verso destra, una macchia informe concentrata prevalentemente sulle alte frequenze (tipica di un rumore di frizione) ①, seguita da una pausa di silenzio ② e da una barra verticale piuttosto stretta ③ (distinta dalle striature verticali che segnalano solitamente la sonorità). Segue poi – appunto – una regione di sonorità ④, caratterizzata da un'ampia striatura orizzontale tipica delle strutture formantiche vocaliche, interrotto nuovamente da una pausa silente (con un'eco parassita)⑤.

Si ripresenta poi un rumore con caratteristiche frequenziali simili a quelle del primo, ma con una distinte dinamiche frequenziali ed energetica ⑥ (questa fase inizia con un'esplosione leggermente accennata). A questo fa seguito una lunga regione di sonorità caratterizzata da un'energia variabile, ma soprattutto da notevoli evoluzioni formantiche.

Dapprincipio, con un graduale aumento d'intensità, si presenta un *pattern* formantico con una certa convergenza in una stessa regione delle formanti più alte (le quali sembrano svilupparsi direttamente dal rumore che segnala il segmento precedente) ⑦. In concomitanza con un aumento d'energia, si ha poi il passaggio a una nuova configurazione abbastanza stabile per un certo tempo ⑧. Questa è a sua volta interrotta da un improvviso calo d'energia associato a una leggera perturbazione del tracciato di evoluzione delle singole formanti (le quali, oltre che essere più deboli, in alcuni tratti sembrano soggette a importanti oscillazioni sulla larghezza di banda) ⑨. L'energia poi riaumenta nel segmento finale che riacquista stabilità nel *pattern* formantico a un'energia comunque relativamente ridotta ⑩.

---

<sup>5</sup> Interessanti contributi in merito a questo delicato tema sono in Abry *et alii* (1985) e Salza (1991) i quali propongono di basarsi su un insieme specifico d'indici temporali per una delimitazione convenzionale dei segmenti.

<sup>6</sup> Sono numerosi nella letteratura internazionale i manuali che si propongono come guide per la lettura degli spettrogrammi (v., tra gli altri, Ferrero *et alii* 1979 e, in ambito internazionale, Ladefoged 1996). Si tratta in alcuni casi di prontuari d'uso (come quello di Painter 1979) che offrono una rassegna d'esempi di realizzazione dei suoni di una determinata lingua studiati spettrograficamente e discussi in termini di tratti caratterizzanti sul piano articolatorio e acustico. Un manuale utile a questo scopo, rivolto all'illustrazione dei suoni dell'italiano, è quello di Giannini & Pettorino (1992) che, oltre a discutere la lettura di numerosi esempi (di laboratorio), propone anche un certo numero di esercizi (con soluzione). In Albano Leoni & Maturi (1995) si trova, invece, un interessante quadro di variazione che espone alcune caratteristiche più critiche della lettura degli spettrogrammi quando relativi a un parlato meno controllato (più spontaneo e, in genere, ipoarticolato).

<sup>7</sup> Possiamo valutare il numero approssimativo di cicli di vibrazione delle corde vocali nell'unità di tempo contando le righe verticali: ad esempio ne rileviamo 8 in 50 ms nel caso della prima vocale (nella finestra delimitata dai primi due demarcatori verticali; in 1 s ce ne saranno 20 volte tante, quindi 160 al secondo) e 12 in 100 ms nel caso dell'ultima vocale (nella finestra delimitata dagli ultimi due demarcatori verticali; in 1 s ce ne saranno 10 volte tante, quindi 120 al secondo). Queste misure indicano una variazione di frequenza fondamentale ( $f_0$ ) da circa 160 Hz a circa 120 Hz, dato tipico di voci maschili (in enunciazioni di tipo dichiarativo).

Ripercorrendo ora queste distinte regioni, proviamo a formulare delle ipotesi sugli indici rilevati per i singoli segmenti, osservando continuità e discontinuità della barra di sonorità (che corre nella parte più bassa dello spettrogramma lungo l'asse del tempo, marcando come sonori i segmenti ④, ⑦, ⑧, ⑨ e ⑩).

Sempre simulando d'ignorare la parola pronunciata, notiamo che il primo segmento ① si candida a rappresentare una fricativa sorda che, date le sue proprietà spettrali (maggiore concentrazione nella banda 5÷8 kHz), non può che corrispondere a un suono alveodentale [s]. Segue la pausa silente ② che contraddistingue la fase di tenuta di un'occlusiva sorda; solo dagli indici (di durata e di estensione frequenziale) dell'esplosione ③ possiamo dedurre che si tratta anche in questo caso di un suono alveodentale [t]. Della vocale seguente ④ possiamo dire che deve trattarsi di una vocale aperta ( $F_1$  è pari all'incirca a 600 Hz o più) e piuttosto centrale o anteriore ( $F_2$  è superiore a 1600 Hz): il fonema vocalico italiano che più spesso presenta realizzazioni con queste caratteristiche è /a/.

Della fase seguente ⑤ possiamo solo dire che rappresenta di nuovo la manifestazione della tenuta di un segmento occlusivo puro o affricato (sordo). Data la presenza di una fase di frizione, preceduta da una regione di esplosione piuttosto dilatata, ⑥ con le proprietà simili a quelle del primo segmento, concludiamo che si tratta qui di un'affricata dentale [ts] (e che, data la sua lunghezza complessiva, si tratti piuttosto di una realizzazione lunga, solitamente rappresentata con [ts̄]).

Il caratteristico schema formantico del segmento ⑦ (con  $F_1$  sotto i 400 Hz e  $F_2$  che corre su valori pari circa a 2300 Hz), unitamente al suo andamento energetico gradualmente crescente, ci fa riconoscere una realizzazione approssimante palatale [j] (una sorta di /i/ consonantica).

Nel segmento ⑧ è facilmente riconoscibile una vocale (che possiamo anche assumere accentata, data la durata della sua fase di maggiore stabilità) che, con i valori approssimativi di 400 Hz per  $F_1$  e 800 Hz per  $F_2$ , è sicuramente alta e posteriore (possiamo pensare a una realizzazione di /o/).

L'ultimo segmento ⑩ è ancora vocalico (debole, ma abbastanza lungo) con  $F_1$  intorno a 500 Hz e  $F_2$  intorno ai 2000 Hz: sebbene arricchito da tracce di nasalità nella struttura formantica, anche per via della sua intensità ridotta, si può riconoscere come una realizzazione di /e/ finale non accentata dell'italiano.

Il segmento tra questi ultimi due, il ⑨, si presenta con un netto calo energetico rispetto a questi: anche per la sua durata e per le caratteristiche formanti 'allargate' (di cui la prima sotto i 400 Hz) si configura tipicamente come una consonante nasale. Per le deviazioni formantiche causate nei segmenti adiacenti, in italiano non può che essere una delle rese possibili di /n/.

Abbiamo così ricomposte una sequenza di tipo [statsjone] che possiamo migliorare, tenendo conto delle durate dei segmenti individuati, trascrivendo [sta'tsjo:ne] (che rappresenta bene una resa non connotata dell'italiano *stazione*, cui – a seconda degli autori – possono corrispondere le forme /sta'tsjone/, /stat'sjone/o /stat'tsjone/).

Le caratteristiche classificatorie su cui basare le considerazioni che portano al riconoscimento dei suoni presenti in una produzione linguistica analizzata su base spettrografica sono state codificate in diverso modo nel corso degli anni. La più celebre tipologia di tratti distintivi basata su considerazioni acustiche è quella proposta da Jakobson, Fant & Halle (1952) e illustrata in Ferrero *et alii* (1979). D'altra parte, diversi approcci di analisi fonologica si sono appoggiati sull'osservazione sperimentale e hanno portato a metodi di parametrizzazione (e di rappresentazione) che, pur limitandosi di solito a poche applicazioni, sono ora diffusamente accolti nei principali trattati di fonetica sperimentale (si veda ad es. Laver 1994).

Si propone in Fig. D un esempio di parametrizzazione articolatoria dell'esempio analizzato in Fig. C che mette in evidenza ampie regioni di sovrapposizione (e trascinamento) dei gesti articolatori necessari per il raggiungimento di determinati obiettivi (bersagli acustici) necessari per la realizzazione del "segmento"<sup>8</sup>.

<sup>8</sup> Si noti che un *tool* di sintesi articolatoria basato sul controllo temporale di una diversa selezione di parametri articolatori (v. Boersma 1998) è disponibile anche in PRAAT (*Artword*).

L'attività (attivazione e/o disattivazione) di un "organo mobile" nel corso del tempo è ovviamente continua, ma può essere valutata in termini binari stabilendo delle soglie convenzionali (presenza, posizione o valore sopra la soglia, vs. assenza, posizione o valore sotto la soglia).

Un flusso d'aria polmonare è presente ad esempio nella realizzazione delle fasi ①, ③, ④, ⑥, ⑦, ⑧, ⑨ e ⑩, ma non in ② e ⑤, caratterizzate da interruzione di questo nella fase di tenuta (occlusione) di [t] e [ts]. In particolare si nota, dalla linea relativa all'attività delle pliche vocali, che solo nelle fasi ④, ⑦, ⑧, ⑨ e ⑩ questo flusso è impiegato per la produzione di energia vocale (sonorità) dato che nelle fasi ①, ③ e ⑥ è invece sfruttato per la produzione di rumore di costrizione, rilascio o esplosione. In tal modo i segmenti [s], [t] e [ts] si caratterizzano come non sonori (sordi).

La parte anteriore della lingua (apice e pre-dorso, il cui abbassamento è controllato in modo prevalente dal muscolo genioglosso) è sicuramente sollevata (oltre una certa soglia, al confine tra suoni vocalici e consonantici) nelle fasi ①, ②, ③, ⑤, ⑥, ⑦ e ⑨, determinando occlusioni, costrizioni o approssimazioni, mentre non lo è nelle fasi (④, ⑧ e ⑩) in cui si determina un'apertura sufficiente alla produzione di suoni vocalici. L'arrotondamento labiale (determinato principalmente dal controllo del muscolo orbicolare) è presente solo in una di queste: la ⑧, relativa alla produzione di [o] che richiede anche un innalzamento e un arretramento del dorso della lingua (regolati dal muscolo stiloglosso, attivo in questa stessa fase). Al contrario, un abbassamento della regione posteriore della lingua (regolato dal muscolo antagonista ioglosso) è invece necessario per gli altri due suoni vocalici (④ e ⑩).

L'abbassamento del velo palatino (controllato dal levatore e dal tensore palatini) è necessario a un certo punto per consentire al flusso di mettere in risonanza le cavità nasali (⑨) per la realizzazione di [n]. Com'è stato messo in evidenza nel grafico, quest'abbassamento incomincia progressivamente prima del previsto e termina con un certo ritardo, determinando parziale nasalizzazione dei segmenti vocalici adiacenti ([o] e, soprattutto, [e])<sup>9</sup>.

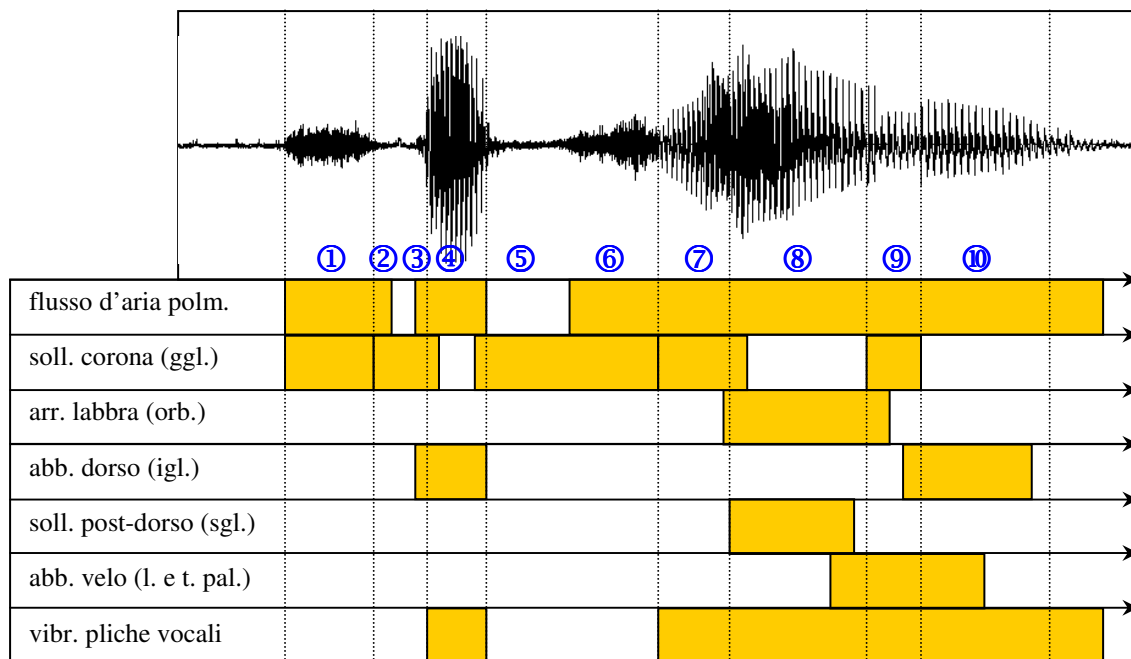


Figura D. Oscillogramma con i segmenti delimitati per l'esempio di Fig. C (in alto) in associazione con gli eventi articolatori verosimilmente responsabili degli indici acustici osservabili nei grafici dell'analisi spettrografica. L'attività di un dato articolatore parametrizzato sull'asse temporale è valutata in termini binari (presenza, in bande colorate, vs. assenza, in bianco; v. testo).

<sup>9</sup> Nel caso di [o], inoltre, dato che quest'abbassamento – in condizioni di sollevamento del dorso – potrebbe portare a un'eccessiva costrizione nella regione dorso-velare, al progressivo sollevamento della corona (necessario per articolare [n] al livello orale) si accompagna una distensione anticipata dello stiloglosso la cui contrazione sarebbe altrimenti in competizione con queste attività.