

Les contraintes inhérentes sont liées à des propriétés de l'organe vibrant, notamment à la force d'inertie des cordes vocales, qui dépend essentiellement de leur masse volumique. Ces facteurs ont une incidence, en particulier, sur la vitesse de changement de F_0 (*speed of pitch change*) et le contrôle de ses trajectoires. Ils donnent ainsi lieu à des phénomènes de troncation de F_0 , lorsque le temps nécessaire à la planification de sa trajectoire est insuffisant, ainsi qu'à des dépassements de la cible à atteindre au terme de cette trajectoire (notion d'*overshoot*). Ces phénomènes feront l'objet d'illustrations dans la section consacrée aux représentations analytiques des paramètres prosodiques. Enfin, les contraintes du troisième type, ou contraintes idiosyncrasiques, sont dépendantes des locuteurs. Elles concernent notamment : (i) les valeurs moyennes du rythme de vibration des cordes vocales (plus rapide chez les femmes, en raison de la plus petite taille des cordes vocales et de leur masse volumique moins importante) et (ii) l'étendue de la tessiture mélodique (ou de la gamme mélodique), qui varie en fonction des capacités vocales des individus.

2. Matérialité acoustique et auditive de la prosodie

Bien que les aspects acoustiques et auditifs de la matérialité prosodique relèvent de niveaux d'analyse distincts, nous avons choisi de les traiter conjointement dans cette partie de l'ouvrage. Le but de ce regroupement est double : commenter d'une part les relations que ces deux niveaux d'analyse entretiennent et examiner, d'autre part, les ruptures de correspondances qui s'instaurent lorsque l'on passe de l'un à l'autre. Avant d'aborder ces deux points, il ne semble pas inutile de rappeler quelques notions élémentaires se rapportant à l'acoustique de la parole.

2.1. Notions élémentaires relatives à l'acoustique de la parole et à la matérialité physique de la prosodie

Le signal acoustique qui transmet la parole est le produit d'une source par une fonction de transfert (ce qui s'exprime par la formule : $P_i = S_i \times T_i$, dans laquelle P désigne le signal produit, S , la source et T la fonction de transfert ; i étant un indice variable). La production de la parole met en jeu deux sources prototypiques : une source de voisement et une source de bruit. La source de voisement, qui est une source sonore périodique entretenue par la vibration des cordes vocales, est mise en jeu dans la production des voyelles et des consonnes dites voisées. Une source de bruit (non périodique) peut se situer en différents lieux du conduit vocal, à la suite d'une fermeture partielle (cas

des consonnes constrictives) ou complète (cas des consonnes occlusives) de ce conduit. La source vocale et les sources de bruit sont à l'origine, séparément ou conjointement, de la production des sons complexes qui forment la parole. La source de voisement est représentée, sur le plan acoustique, par un spectre de raies, qui comporte un fondamental et un certain nombre d'harmoniques dont les valeurs sont des multiples entiers de celle du fondamental. Par exemple, si la valeur du fondamental est de 150 Hz, celle du premier harmonique sera de 300 Hz, celle du second harmonique, de 450 Hz, etc. Le spectre d'une source de bruit ne contient pas de fondamental, mais un nombre théoriquement infini de fréquences, qui ne sont pas dans un rapport harmonique comme c'est le cas pour la source vocale périodique. Le conduit vocal, qui s'étend de la glotte aux lèvres, agit comme un filtre acoustique en sélectionnant certaines fréquences du spectre de la source (périodique ou non périodique), qui bénéficient ainsi d'un transfert d'énergie (d'où l'expression fonction de transfert). Les régions du spectre de la source sélectionnées de la sorte correspondent à des maxima d'amplitude du spectre du son produit à la sortie du conduit vocal. Ces maxima sont qualifiés de formants dans le cas des voyelles et de pôles de bruit, dans celui des consonnes non voisées.

Les valeurs de la fréquence fondamentale (F0) d'un signal, qui s'expriment en Hertz (Hz), varient en raison inverse de celles de la périodicité de ce signal (ce qui se traduit par la formule : $F = 1/T$, dans laquelle F désigne la fréquence et T la période). La fréquence fondamentale est analysée auditivement par l'être humain sur la base d'un calcul qui prend en compte l'intervalle entre les harmoniques dans le bas du spectre. Ceci explique pourquoi la fréquence fondamentale peut être détectée, même en l'absence physique de la composante fondamentale. Les variations de F0 sont analysées auditivement comme des changements de hauteur ou de mélodie (cf. figure 10). Les termes de F0 et de mélodie ne sont donc pas interchangeable, puisqu'ils renvoient à des niveaux d'analyse différents (cf. le chapitre 7, sur les niveaux d'analyse de la prosodie). Les écarts entre les différentes valeurs de F0 d'un signal peuvent être calculés au moyen d'échelles prétendues auditives, qui expriment ces rapports en octaves, fractions d'octaves ou demi-tons.

Les variations du spectre acoustique des sons sont analysées auditivement comme des changements de timbre, ce dernier pouvant être qualifié subjectivement, par exemple, de grave ou d'aigu, de riche ou de pauvre, etc. C'est la disposition et l'amplitude relatives des formants et des pôles de bruits dans le spectre qui déterminent, respectivement, le timbre des sons vocaliques et celui des consonnes. Les voyelles disposent ainsi subjectivement d'une hauteur et d'un timbre, alors que les consonnes peuvent posséder un timbre (on peut distinguer, en effet, entre des consonnes graves et des consonnes aiguës), mais n'ont pas de hauteur à proprement parler. Il est donc logique que le rôle des voyelles se révèle déterminant pour l'analyse prosodique, en général et pour

celui de l'intonation, en particulier. Le spectre (et donc le timbre) a aussi un rôle à jouer en prosodie, notamment en ce qui concerne l'analyse des phénomènes accentuels. En effet, les voyelles accentuées, qui sont émises avec une force articulatoire plus grande, disposent généralement d'un spectre acoustique mieux défini que celui des voyelles inaccentuées. Ces dernières sont sensibles à un effet de réduction, qui est plus ou moins accusé selon la langue. Par exemple, le phénomène est beaucoup plus marqué en anglais qu'en français, la réduction pouvant aller jusqu'à la neutralisation du timbre des voyelles en anglais, qui tend ainsi à se rapprocher de celui de la voyelle neutre (ou *schwa*).

La durée physique des unités qui constituent le signal de la parole et qui se mesure conventionnellement en millisecondes (ms), constitue également un paramètre dont l'importance est cruciale pour les analyses prosodiques. Toutefois, la représentation analytique de ce paramètre ne peut faire l'objet d'une procédure directe, comme c'est le cas pour F0 ou pour l'intensité, car cette représentation nécessite un choix préalable de l'unité sur laquelle on souhaite faire porter la mesure. Il n'est pas possible en effet, comme nous l'avons déjà rappelé, d'obtenir directement une courbe des variations de la durée, contrairement à ce que l'on obtient pour des paramètres comme la F0 ou l'intensité. Selon les objectifs de la recherche, l'unité retenue pour effectuer les mesures de durée pourra être la phrase, des membres de phrases, le mot, la syllabe, ou des sons individuels, voire des portions de sons, comme l'attaque, le noyau et la coda des segments vocaliques.

Dans les travaux sur la prosodie, l'intensité est un paramètre qui n'a pas fait l'objet de soins aussi attentifs que ceux qui sont accordés à la F0 et à la durée, probablement en raison des difficultés inhérentes à sa mesure physique et à son évaluation subjective. En réalité, le concept d'intensité fait référence à plusieurs notions, telles que : la pression sonore, la puissance, le niveau sonore, le volume sonore, l'énergie, etc., dont les liens ne sont pas toujours évidents pour le profane. De surcroît, il convient de distinguer entre l'intensité objective (*intensity* en anglais) et l'intensité subjective (*loudness* en anglais), ainsi qu'entre les diverses unités de mesure et les échelles correspondantes. Enfin, l'estimation de l'intensité subjective, pour les sons de la parole continue, représente une tâche particulièrement difficile, en raison du nombre de facteurs et de précautions qui doivent être pris en compte. Nous éviterons d'entrer ici dans des précisions que l'on pourra trouver dans les ouvrages spécialisés et nous nous limiterons donc à exposer les points qui nous paraissent essentiels.

À l'instar des autres signaux sonores, les signaux physiques par lesquels se transmet la parole exhibent, lorsqu'on visualise leurs tracés, des variations de l'amplitude en fonction du temps. Ces changements d'amplitude correspondent en fait à des variations de la pression acoustique, qui s'exercent d'abord sur le tympan et qui font ensuite l'objet de divers traitements par le système auditif. Le niveau de la pression acoustique que capte l'organe auditif et qui détermine ainsi la puissance sonore d'écoute, est généralement qualifié

de niveau de pression sonore, lequel est conventionnellement représenté par le sigle SPL (pour *sound pressure level*). Il est possible d'utiliser deux grandeurs pour calculer l'intensité sonore objective : la pression, qui se mesure en pascals (pa) et l'intensité, qui se mesure en watts par cm^2 (soit en termes de puissance par unité de surface). Cependant, l'approche commune de l'intensité sonore se réfère à une échelle unique : l'échelle des décibels (dB). Il s'agit d'une échelle logarithmique, qui exprime les différences d'intensité par des rapports, en fonction d'une valeur de référence, définie conventionnellement comme le 0 dB. Cette référence correspond au seuil d'audibilité pour un son dont la fréquence est égale à 1Kz (soit : $2 \cdot 10^{-5}$ pa, pour la pression et 10^{-16} Watt/ cm^2 , pour l'intensité). Si le seuil d'audibilité est de 0 dB, le seuil de tolérance, ou seuil de la douleur, est équivalent à 130/140 dB environ, ce qui représente un écart considérable et une gamme de valeurs très étendue, que seule une échelle logarithmique permet, en vérité, de manipuler sans trop de difficultés. De surcroît, cette échelle correspond davantage à ce que capte l'oreille humaine en termes de sensation. Le logarithme d'un nombre est l'exposant qu'il faut attribuer à la base (ici la base = 10) pour obtenir ce nombre. Par exemple :

- 10 : le logarithme est 1, car $10^1 = 10$
- 100 : le logarithme est 2, car $10^2 = 100$
- 1000 : le logarithme est 3, car $10^3 = 1000$

Étant donné que la valeur de l'intensité en dB (I) d'un son donné se calcule par la formule : $I \text{ dB} = 10 \log_{10} I/I_0$ (I_0 étant la valeur de référence), on obtient les échelles de correspondances suivantes.

Echelle					
	0	10	100	1000	10 000
		10^1	10^2	10^3	10^4
log. base 10		1	2	3	4
10 log. base 10		10	20	30	40
					décibels (dB)

Tableau II. Échelle logarithmique utilisée pour le calcul des décibels.

Le calcul de l'intensité objective des sons de la parole utilise l'échelle des décibels. Mais l'intensité se mesure habituellement en valeurs RMS (de l'anglais *root mean square*, ou moyenne quadratique), dites aussi valeurs efficaces, qui se basent sur le calcul de l'amplitude continue équivalente en puissance (figure 13). Le calcul tient également compte de l'intégration temporelle du signal en adoptant une constante de temps d'intégration (ct) déterminée. C'est ainsi, par exemple, que des (ct) de 5 ms et de 10 ms signifient que le signal est analysé toutes les 5 ms et 10 ms, respectivement.

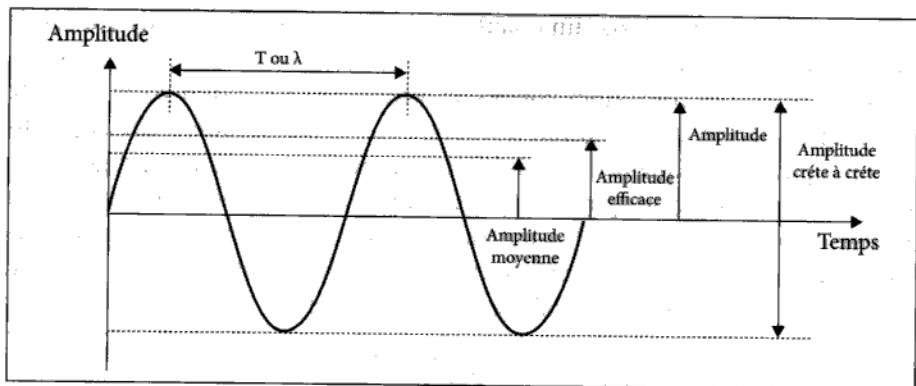


Figure 13. Illustration des différentes façons de définir de l'amplitude d'un signal acoustique. T désigne la période de ce signal.

2.2. Ruptures des correspondances entre la matérialité acoustique et la matérialité auditive de la prosodie

Entre les propriétés physiques des paramètres prosodiques et leur évaluation subjective, on relève d'importantes différences, qui sont dues à des contraintes de bas niveau imposées par l'organe auditif et à des effets de normalisation de niveau plus élevé, qui sont attribuables au système de perception. Nous commencerons par examiner l'exemple de F0, qui appelle le plus de remarques sur ce point. Nous avons déjà mentionné, dans le cours de cet ouvrage, que l'évolution de F0 en fonction du temps est représentée par un tracé qui présente des interruptions et des perturbations locales. Les premières correspondent à la réalisation des consonnes non voisées et les secondes à des interférences entre la production des diverses consonnes et le suivi de F0. Ces perturbations locales sont illustrées par la figure 14, où il apparaît, d'une part, que les consonnes non voisées provoquent un saut de F0 (*pitch skip* en anglais) à l'initiale des voyelles subséquentes et, d'autre part, que les consonnes voisées sont marquées par des creux de F0 plus ou moins importants. En règle générale, il s'agit de sauts et de creux plus accusés pour les consonnes obstruantes voisées que pour les consonnes sonorantes comme [l, m, n]. Ces variations sont associées à des effets aérodynamiques, liés au jeu de glissement entre la source sonore et le conduit vocal (cf. *supra*). Pour les consonnes non voisées, la glotte est ouverte. L'air pulmonaire traverse cette ouverture provoque une augmentation de la pression dans le conduit vocal, en amont du point d'articulation. À la fin de la consonne qui précède le début de la voyelle, la chute de la pression intra-glottique, entraîne momentanément une forte augmentation de la pression trans-glottique, lequel provoque un sursaut vibratoire (le sursaut de la glotte en position de phonation) pendant un très court