

## 2

## Mécanismes physiques de la voix

La parole est produite lorsqu'un locuteur génère un champ sonore au sein de son appareil vocal. Le champ sonore se propage ensuite à travers le conduit vocal qui est constitué de l'ensemble des cavités comprises entre le larynx, les lèvres et les narines. Sa forme exacte dépend de la position des articulatoires que sont les lèvres, les dents, la mâchoire, la langue et le voile du palais. Des formes distinctes du conduit vocal modifient le champ sonore qui s'y propage, produisant ainsi les différents sons de parole.

Un champ sonore est créé lorsque le débit de l'air expiratoire est obligé à varier rapidement. La source sonore qui est de loin la plus versatile, la plus fréquemment utilisée, la plus complexe et la plus étudiée est formée par l'écoulement pulsé de l'air entre les cordes vocales qui vibrent. Les cordes vocales sont constituées de deux plis charnus qui sont tendus en travers du larynx, à la jonction entre la trachée et le pharynx. Le terme anatomique est en fait « plis vocaux », conformément à la terminologie des *Nomina Anatomica* : « *plicae vocales* ». Les plis vocaux participent également à la fonction sphinctérienne du larynx, en dehors du contexte de la communication orale ou du chant.

Le son qui est généré à l'aide de la vibration des cordes vocales est appelé voix et les sons de parole qui sont produits en propageant la voix à travers le conduit vocal sont dits voisés. En français, les sons non voisés sont les plosives et les fricatives qui se trouvent au début des mots comme : pas, tas, cas et fa, sa, chat. Les affriquées non-voisées sont des sons composites qui se trouvent au début de quelques mots importés, tels que tsar et tchador. Tous les autres sons du français sont voisés par défaut.

Les pathologies du larynx, dans la mesure où elles affectent la vibration des plis vocaux, causent les troubles de la voix qui sont généralement les effets perceptibles d'une anomalie laryngée. La longueur de la chaîne de causalité qui lie pathologies laryngées et timbres vocaux perçus, l'observation difficile des mouvements des plis vocaux, mais aussi l'importance de la voix en tant que telle dans la communication orale et le chant ont incité les laryngologues et autres experts à décrire les troubles de la voix en rapportant des propriétés de la parole émise ou des propriétés d'autres signaux obtenus non-invasivement qui renseignent sur les vibrations laryngées. Idéalement, les

chercheurs préfèrent étudier des propriétés du signal de parole qui constituent des indices perceptivement pertinents, c'est-à-dire des indices qui prédisent la qualité vocale perçue.

Ce qui suit tente d'expliquer qualitativement le maintien de la vibration des cordes vocales en invoquant une chaîne de causalité circulaire entre le flux d'air respiratoire et l'obstacle que forment les plis laryngés sur son passage. Suit une discussion des déviations des cycles vocaux de la périodicité stricte, déviations qui sont des indices sonores du degré d'enrouement perçu. L'essentiel du chapitre concerne une classification possible de l'ensemble des indices sonores proposés afin de décrire les anomalies de la voix. Ces propositions sont multiples et disparates dans leurs objectifs. Le chapitre se termine par une brève discussion des problèmes connus des indices actuels du timbre vocal.

## **Voix et vibrations des plis vocaux**

Les cordes vocales désignent deux plis charnus composites situés à l'intérieur du larynx à la jonction de la trachée et du pharynx. Ces plis sont orientés dans le sens antérieur-postérieur. Ils s'appuient à l'avant sur le cartilage thyroïde (la pomme d'Adam), et vers l'arrière sur deux petits cartilages appelés aryténoïdes. Le locuteur contrôle l'écartement, la longueur, la forme et la tension des plis vocaux en contractant le muscle vocal ou en changeant les positions des cartilages du larynx les uns par rapport aux autres. Le volume tridimensionnel compris entre les plis vocaux est appelé glotte (Shearer, 1963 ; Sundberg, 1987).

Le son produit par l'écoulement pulsé de l'air à travers la glotte, qui change de forme et de taille, est appelé voix. En régulant la vibration des plis vocaux, le locuteur fixe le timbre, la hauteur et la sonie de son émission vocale, contrôle l'intonation des phrases, accentue les syllabes et bascule entre le voisement et le non-voisement des sons de parole (Ladefoged, 2005). La production de la voix exige un débit d'air pulsé. La pulsation du débit d'air est causée par la vibration des plis vocaux. En général, la production de la voix implique la répétition d'un même cycle d'ouverture et de fermeture de la glotte à raison d'une centaine de fois à la seconde.

La vibration des plis vocaux est auto-entretenu. Le mouvement des plis détermine la taille des forces qui causent le mouvement et ainsi de suite. Les forces sont de trois types. Les premières sont les forces dues à l'élasticité des plis. Les forces élastiques s'opposent à la déformation et au déplacement des plis de leur position d'équilibre et favorisent le retour à la position de départ des plis déplacés. Les deuxièmes sont les forces de freinages. L'amortissement du mouvement des plis est dû à la viscosité des plis et aux collisions du pli gauche et droit au moment de la fermeture de la glotte.

Les forces de freinages s'opposent toujours au mouvement. Afin de maintenir le mouvement, un apport continu d'énergie de l'extérieur est nécessaire. Cette énergie est apportée par le flux d'air pulmonaire (Giovanni et coll., 2002). Les troisièmes sont les forces qui sont dues à la pression de l'air dans la glotte. Cette pression dépend de la pression de l'air dans les poumons, dans la trachée ainsi que de la vitesse variable de l'écoulement de l'air à travers la glotte dont la taille et la forme évoluent.

Deux autres éléments expliquent la vibration. Le premier concerne les masses des plis vocaux. Les masses confèrent aux plis une inertie qui maintient leur mouvement en l'absence d'autres forces. Le deuxième est la forme changeante de la glotte car l'air s'écoule plus rapidement dans une glotte divergente que convergente. Une glotte divergente est une glotte qui est plus large en aval qu'en amont. Le changement de la vitesse de l'écoulement au cours d'un cycle de vibration est crucial pour le maintien des vibrations car un flux d'air rapide exerce une force plus faible sur les plis qu'un flux lent (Titze, 1994 ; Alipour et Scherer, 2000).

À ce stade sont réunis tous les éléments nécessaires à une explication qualitative de la vibration (Ishizaka et Flanagan, 1972 ; Gunter, 2003). Le locuteur prépare le voisement en tendant les plis et en pivotant les cartilages aryténoïdes qui rapprochent les plis du plan sagittal médian du larynx. À ce moment, la glotte est fermée ou presque fermée. Le locuteur démarre la vibration en augmentant la pression de l'air dans la trachée. La pression dans la trachée va vaincre les forces élastiques des plis qui finissent par s'écarter par en bas. La forme de la glotte est convergente c'est-à-dire que la glotte est plus large en amont qu'en aval (étape a). Une fois que les plis commencent à s'écarter, le flux d'air démarre à travers la glotte. Il continue à écarter les plis aussi longtemps que les forces aérodynamiques sont plus grandes que les forces élastiques. En général, le mouvement d'écartement se poursuit jusqu'à ce que le bord supérieur de la glotte (la sortie du côté du pharynx) soit aussi large que le bord inférieur (l'entrée du côté de la trachée). À ce moment, la forme de la glotte change d'une forme convergente en une forme divergente (étape b). Dès que la forme est devenue divergente, le jet d'air décolle de la paroi de la glotte à l'endroit où la section droite est la plus faible. Par conséquent, la vitesse de l'écoulement augmente et la pression de l'air à l'intérieur de la glotte diminue. Les forces exercées par l'air sur les plis ne peuvent plus contrecarrer les forces élastiques, et la glotte se ferme. Au moment de la fermeture complète, tout mouvement s'arrête, et le processus reprend à l'étape a (étape c).

Le voisement s'arrête lorsque la pression dans la trachée diminue trop parce que l'air pulmonaire est épuisé et que le locuteur doit inspirer, ou plus normalement, lorsque le locuteur écarte les plis en pivotant les aryténoïdes vers l'extérieur. L'écartement des plis rend impossible le maintien de la pression de l'air dans la trachée et les vibrations s'arrêtent.

L'amorçage des vibrations requiert que la pression de l'air sous la glotte dépasse un seuil critique à cause de l'inertie des plis et leur viscosité. Une fois la phonation entrée en régime, la pression sous-glottique peut descendre sous ce seuil (Lucero, 1999). Cette pression-seuil a retenu l'attention des cliniciens car elle dépend de la viscosité de la muqueuse des cordes vocales (Titze et Schmitt, 1994 ; Elliot et coll., 1995 ; Hemler et coll., 1997 ; Vilkman et coll., 1999 ; Jiang et coll., 2001).

## **Troubles de la voix et irrégularités des cycles vocaux**

Les troubles de la voix sont les conséquences audibles d'une malformation organique acquise ou congénitale des plis vocaux ou d'un contrôle déficient de la respiration ou de l'articulation laryngée. Les malformations organiques acquises peuvent être la conséquence d'un malmenage ou surmenage vocal.

Des symptômes fréquents des troubles de la voix sont des irrégularités des cycles de parole. Les origines de ces irrégularités sont multiples.

### **Régimes de vibrations non-modaux**

L'ensemble des plis vocaux et du flux d'air glottique forme un système dynamique non linéaire qui peut exhiber de multiples régimes vibratoires qui sont qualitativement différents. Des exemples de régimes vibratoires distincts sont les mécanismes laryngés qui sont à la base des registres de parole. On différencie généralement trois registres de parole qui sont le craqué, le modal et le falsetto. Ils sont caractérisés par des cycles de parole réguliers, c'est-à-dire que les cycles successifs diffèrent peu en durée, en forme ou en amplitude.

Les plis vocaux peuvent vibrer dans des régimes caractérisés par des cycles adjacents différents. Les sons de parole sont, à ce moment, souvent perçus comme enroués. On peut distinguer les régimes suivants, sur la base des spectres des sons de parole (Behrman et coll., 1998) :

- le régime dit modal ou neutre produit des cycles de parole réguliers avec un spectre qui comprend une seule série d'harmoniques ;
- les voix diplophoniques sont caractérisées par de brèves séquences de cycles vocaux inégaux qui se répètent. Les spectres comprennent plusieurs séries d'harmoniques dont les fréquences fondamentales sont dans des rapports rationnels simples ;
- la bi-phonation désigne un régime vibratoire qui est caractérisé par un spectre discret comprenant des séries d'harmoniques dont les fréquences fondamentales sont dans des rapports irrationnels. Par conséquent, le signal de parole est a périodique ;

- certaines voix sont caractérisées par des séquences de cycles dont les durées, les amplitudes ou les formes fluctuent aléatoirement. Les spectres correspondants sont continus, la voix est perçue comme rauque.

### **Amplification de la gigue et du *shimmy* vocal**

La gigue vocale (*vocal jitter*) désigne les petites perturbations rapides des durées des cycles glottiques. Les origines sont inconnues. On cite couramment des causes neurologiques, l'écoulement turbulent de l'air à travers la glotte, la répartition inégale de mucus sur les plis vocaux... (Pinto et Titze, 1990). L'amplification de la gigue vocale est parfois observée en présence de pathologies ; elle devrait s'expliquer par une modification de la réponse du système laryngé à des perturbations qui, en tant que telles, ne sont pas nécessairement affectées par la pathologie (Schoentgen, 2001).

Le *shimmy* vocal (*vocal shimmy* ou *shimmer*) désigne de petites perturbations de l'amplitude des cycles de parole. Les causes du *shimmy* vocal sont supposées être les mêmes que celles de la gigue vocale. Mais la transformation dans le conduit vocal de la gigue vocale en *shimmy* vocal, ainsi que le report variable d'énergie acoustique d'un cycle à l'autre contribuent également au *shimmy* des cycles de parole (Schoentgen, 2003).

Souvent, une petite perturbation rapide des durées et des amplitudes des cycles de parole est désignée par les termes anglais *jitter* et *shimmer* (ou *shimmy*) qu'elle qu'en soit la cause exacte.

### **Tremblement vocal**

Le tremblement vocal désigne des oscillations lentes de la fréquence (1-15 Hz) ou de l'amplitude instantanée des sons de parole voisés. Les causes sont physiologiques (respiration, battement cardiaque) et pathologiques (parkinsonisme, tremblement cérébelleux...). Par analogie avec le *shimmy* vocal, le tremblement de la fréquence vocale contribue au tremblement de l'amplitude des cycles de parole. Une cause extra-laryngée est le tremblement des articulateurs du conduit vocal (Dejonckere et coll., 1996 ; Schoentgen, 2002 et 2003).

### **Bruit de turbulence excessif**

L'écoulement turbulent de l'air à travers un orifice ou la collision d'un jet d'air turbulent avec un obstacle génère un signal acoustique. Il est perçu comme un bruit à large bande spectrale quand la turbulence est pleinement développée. L'intensité du bruit dépend de l'interaction du flux avec des obstacles éventuels, ainsi que de la vitesse de l'écoulement (Stevens, 1998).

En présence de pathologies, l'intensité de ce bruit peut augmenter et donner lieu à du souffle audible. La voix soufflée est souvent assimilée à une voix faible, à cause de la fermeture brève ou incomplète de la glotte et le débit d'air accru qui caractérise les voix basses. Ce lien n'est pas inéluctable (Isshiki et coll., 1978 ; Hillman et coll., 1983 ; Wolfe et Martin, 1997).

### **Vibrations non sollicitées de structures du larynx qui normalement ne vibrent pas**

Le larynx comprend, outre des plis vocaux, deux paires de plis supplémentaires qui normalement ne vibrent pas. Ces plis sont les bandes ventriculaires (ou plis vestibulaires) et les plis ary-épiglottiques. Ils sont mis en vibration occasionnellement à des fins artistiques ou phonétiques (Essling, 1999). Mais, ils peuvent vibrer à l'insu du locuteur suite à un contrôle fautif de la position des cartilages du larynx ou des tensions des muscles laryngés. La voix qui implique une vibration des bandes ventriculaires est occasionnellement appelée une voix de bandes. La voix de bandes est perçue comme rauque.

### **Transitions incontrôlées entre différents régimes de vibration**

Des locuteurs dysphoniques peuvent basculer de façon incontrôlée d'un registre de parole à un autre, d'un régime régulier à un régime qui l'est moins, ou subir des arrêts momentanés du voisement. Parfois, ces transitions sont appelées des bifurcations, par analogie avec les systèmes dynamiques non linéaires. Ils peuvent fonctionner dans des régimes qualitativement différents. Le passage d'un régime à un autre se fait subitement suite au franchissement d'un seuil critique par un des paramètres du système dynamique (Berry et coll., 1996 ; Spencer et Titze, 2001).

### **Indices phonatoires des troubles de la voix**

Cette section donne une vue d'ensemble des indices phonatoires des troubles de la voix et propose un cadre de classification. En général, les indices phonatoires ou indices acoustiques sont des attributs numériques ou mesures qui résument des propriétés du signal de parole, ou d'autres signaux qui ont été obtenus de façon non invasive, et qui ont une pertinence clinique. Typiquement, l'acquisition d'indices phonatoires implique l'enregistrement (de façon non invasive) de signaux qui renseignent sur le fonctionnement laryngé, le traitement de signal qui évacue des propriétés superflues du signal et le résumé des propriétés cliniquement pertinentes à l'aide de quelques nombres.

Les objectifs de l'extraction d'indices phonatoires dans un contexte clinique sont la documentation de la voix des patients, le suivi longitudinal pendant le traitement (avant et après chirurgie, par exemple) ainsi que la comparaison avec des locuteurs normophoniques.

Présenter une classification des formes mathématiques en tant que telles des indices phonatoires serait une tentative vaine car la grande majorité des indices existants ont été définis heuristiquement ; leurs propriétés mathématiques et statistiques n'ont été étudiées que superficiellement. L'objectif est de classer les indices phonatoires en rapport avec leur usage.

En effet, la littérature scientifique et clinique consacrée aux indices acoustiques des troubles de la voix est abondante. Malheureusement, sa diversité est impressionnante. Elle est liée aux handicaps et pathologies étudiés, aux symptômes vocaux décrits, aux tâches vocales demandées aux locuteurs, aux performances extra-linguistiques, para-linguistiques et linguistiques examinées, aux capteurs utilisés, aux matériaux sonores enregistrés, ainsi qu'aux corrélations étudiées.

Par conséquent, il est parfois difficile de comparer des résultats obtenus dans des contextes différents et d'en tirer des règles générales.

## **Classification**

La raison d'être d'une classification est de faciliter la planification de nouvelles expériences et l'exploitation de la littérature existante. Suit ici une proposition de grille de classification de différentes approches de l'évaluation des troubles de la voix, fondées sur des indices acoustiques.

### **Étiologie**

Une grande variété de pathologies et handicaps qui ont un effet sur la voix a été examinée. Typiquement, sont distingués les troubles de la voix résultant de changements organiques des plis vocaux et les troubles dits dysfonctionnels qui ne sont pas la conséquence de changements structuraux ou organiques directement observables, mais éventuellement, la conséquence d'un contrôle non optimal de la part du locuteur (Boone et coll., 2004). Les troubles de la voix causés par des problèmes moteurs constituent une troisième catégorie. Les voix de locuteurs parkinsoniens font partie de cette catégorie, par exemple. Les voix dites de substitution forment une catégorie à part. Ce sont des voix dont la production repose sur des mécanismes artificiels ou naturels alternatifs. Elles permettent à des personnes qui ont perdu la faculté de produire la voix grâce aux vrais plis vocaux, de communiquer oralement.

La question est de savoir si des troubles de la voix qui ont des causes différentes devraient être décrits à l'aide d'indices phonatoires distincts ; cette question est restée sans réponse. En général, des problèmes de la parole et de la voix qui sont causés par des troubles moteurs sont examinés séparément des

troubles causés par des pathologies laryngées (Till et coll., 1994 ; Bunton et Weismer, 2002). Une exception possible est la paralysie des cordes vocales.

### ***Matériaux linguistiques***

Une distinction majeure, entre ces différentes approches de l'évaluation vocale, repose sur le matériel phonétique qui est examiné. En effet, le matériel analysé peut être de la parole connectée ou des sons voisés soutenus (Yu et coll., 2002). Ces derniers peuvent à leur tour être subdivisés selon que les attaques et les déclins sont inclus ou non dans l'intervalle d'analyse (Révis et coll., 2000).

L'analyse de la parole connectée est souvent présentée comme un idéal. L'analyse des voyelles soutenues est fréquente (Giovanni et coll., 1999 ; Yu et coll., 2001 ; Bhuta et coll., 2004). Beaucoup de méthodes d'analyses reposent sur l'hypothèse de la stationnarité et périodicité locale du signal. Ces hypothèses ne sont pas vérifiées pour des locuteurs sévèrement enrôlés ou des locuteurs produisant de la parole connectée.

### ***Tâches***

Les tâches se rapportent aux activités qui sont exigées des locuteurs pendant l'évaluation vocale. Les tâches les plus courantes sont la production de la parole, y compris le soutien de sons de parole isolés (Dejonckere et coll., 1996), le chant (éventuellement) (Elliot et coll., 1995 ; Rothman et coll., 2001), le chargement vocal ainsi que le profilage vocal.

Le chargement vocal consiste en la mesure des indices phonatoires d'un locuteur, suivi par la lecture à haute voix pendant un laps de temps important (pendant 45 minutes, par exemple) puis d'une nouvelle mesure des mêmes indices. L'objectif est de suivre des modifications vocales qui seraient la conséquence d'une charge vocale importante (Vilkman et coll., 1997 ; Mann et coll., 1999 ; Rantala et Vilkman, 1999 ; Vilkman et coll., 1999 ; Bäckstrom et coll., 2003).

Le profilage est la mise en évidence des limites des performances vocales d'un sujet, c'est-à-dire la voix la plus faible et la plus forte possible, la voix la plus aiguë et la plus grave possible, et ainsi de suite. Ces limites ne sont pas toujours absolues. Elles peuvent être déterminées par rapport à une tâche, la lecture d'un texte à haute voix, par exemple (Stemple, 1993 ; Heylen et coll., 2002). Le temps maximum de phonation explore les limites des performances vocales, mais il est souvent assimilé à une mesure des performances aérodynamiques.

### ***Performances***

Les performances renvoient aux capacités vocales qui sont examinées. Les performances phonatoires peuvent être subdivisées en registres, types phonatoires, voisement, prosodie et qualité vocale (timbre).

Les registres de parole qui sont couramment décrits sont, du plus grave au plus aigu, le craqué, la voix modale et le falsetto (Spencer et Titze, 2001 ; Heylen et coll., 2002).

Les types phonatoires sont situés dans un intervalle qui embrasse l'étendue entre la phonation hypo-fonctionnelle et hyper-fonctionnelle. Couramment, une voix faible ou soufflée est assimilée à une phonation hypo-fonctionnelle et une voix pressée à une phonation hyper-fonctionnelle (Alku et Vilkman, 1996 ; Vilkman et coll., 2002 ; Bergan et coll., 2004). Les types phonatoires intermédiaires sont appelés neutres ou normaux et, occasionnellement, fluides (*flow phonation*). Le terme « voix faible » serait plus approprié que « voix soufflée », car la perception éventuelle d'un excès de bruit de turbulence n'est pas réservée aux voix hypo-fonctionnelles.

Le voisement renvoie à la faculté du locuteur de basculer entre la vibration et l'absence de vibration des plis vocaux lors de la production des sons de parole afin de réaliser l'opposition phonétique entre segments voisés et non-voisés (Bunton et Weismer, 2002). La prosodie repose sur la faculté du locuteur de contrôler l'intonation, l'accentuation et le rythme ainsi que le débit de parole (Coleman et Markham, 1991 ; Stone et Rainey, 1991). Finalement, la qualité vocale désigne le timbre vocal, c'est-à-dire l'enrouement, la raucité, le chevrottement, le tremblement vocal... (Wolfé et Martin, 1997 ; Schoentgen et coll., 2000 ; Wuyts et coll., 2000).

### **Capteurs**

Les capteurs sont utilisés afin d'enregistrer de façon non-invasive des signaux qui rapportent sur les performances phonatoires du locuteur. Le signal microphonique est utilisé le plus souvent. Il évolue proportionnellement à la pression acoustique et, par conséquent, proportionnellement au signal de parole qui est perçu par l'auditeur. Le microphone de contact capte les vibrations acoustiques au niveau du cou (Ohlsson et coll., 1989). D'autres signaux qui peuvent être obtenus de façon non-invasive sont l'électroglottogramme et le photoglottogramme (Laukkanen et coll., 1999 ; Habermann et coll., 2000). Le premier évolue proportionnellement à la surface de contact entre les cordes vocales et le deuxième proportionnellement à la section droite de la glotte. Les signaux aérodynamiques sont la pression intra-orale, le débit oral et le débit nasal captés à l'aide d'un masque (Giovanni et coll., 1999 ; Vilkman et coll., 1999 ; Morsomme et coll., 2001 ; Vilkman et coll., 2002 ; Yu et coll., 2002).

### **Signal d'excitation glottique versus signal de parole**

Les indices acoustiques des troubles de la voix sont extraits à partir du signal phonatoire ou à partir du signal de parole. Le signal d'excitation glottique est le signal acoustique qui est généré à la glotte par la vibration des plis vocaux et le flux d'air pulsé. La parole voisée est émise dans le plan des

lèvres suite à la propagation acoustique du signal d'excitation glottique à travers le conduit vocal. L'observation directe du signal d'excitation glottique ou son inférence indirecte à partir du signal de parole est difficile (Vilkman et coll., 1997 ; Alku et coll., 1998 ; Alku et coll., 2001). Souvent, il est remplacé par des signaux auxiliaires comme l'électroglottogramme ou le photoglottogramme qui informent sur des propriétés de la glotte directement.

### ***Domaines de transformation***

À présent, une classification systématique des méthodes de traitement du signal de parole à vocation clinique n'est pas possible, car la majorité des méthodes comprennent des étapes heuristiques de traitement qui diffèrent d'une tâche à l'autre et même d'une étude à l'autre.

Une classification possible qui est liée au traitement du signal est basée sur les transformations mathématiques qui sont utilisées en option. Ces transformations modifient le caractère du signal afin de faciliter la description des symptômes vocaux. Des exemples sont la transformation de Fourier, de Hilbert ou la transformation en ondelettes (Alku et coll., 1998 ; Schoentgen et coll., 2000 ; Bäckstrom et coll., 2003). Si aucune transformation n'est appliquée, alors les indices phonatoires sont temporels.

### ***Symptômes vocaux***

Une distinction essentielle entre indices repose sur les symptômes vocaux qui sont visés. Les symptômes vocaux sont les propriétés de la parole qui sont cliniquement pertinentes, qui sont affectées par l'état de glotte et qu'on tente de mettre en évidence à l'aide d'un traitement du signal. Typiquement, on fait la distinction entre les dyspériodicités du signal, la morphologie du signal, les traits supra-segmentaux, et la coordination entre événements glottiques et supraglottiques.

La coordination se réfère au *timing* des débuts et fins du voisement par rapport à des événements supraglottiques. Des exemples d'événements supraglottiques pertinents sont la fermeture et la détente des consonnes ou l'attaque et le déclin des voyelles. Un trait de coordination qui est souvent étudié est le délai d'établissement du voisement qui est l'intervalle temporel entre la détente d'une consonne plosive ou fricative et le début du voisement de la voyelle qui suit. Cet intervalle est positif lorsque le voisement suit la détente et négatif dans le cas opposé. Le délai d'établissement du voisement est un trait perceptif sur lequel repose la distinction entre consonnes voisées et non-voisées. Cet intervalle, qui dépend de la langue, peut être très bref et exige un contrôle fin du *timing* de l'initiation du voisement par rapport à la détente de la consonne. C'est pourquoi, le délai d'établissement du voisement est souvent étudié dans le contexte des troubles moteurs ou des voix de substitution qui sont susceptibles de rendre cette coordination difficile (Bunton et Weismer, 2002).

Les traits supra-segmentaux renvoient à l'intonation, à l'accentuation et aux tons. Le débit de parole (en nombre de syllabes par unité de temps), la fréquence phonatoire moyenne, son écart-type ainsi que le niveau de pression sonore (en dB) sont aussi des traits supra-segmentaux, mais leur rôle est plutôt stylistique que linguistique. En clinique, le débit de parole, la moyenne et la variabilité ou l'étendue de la fréquence phonatoire, et le niveau de pression sonore sont souvent étudiés (Coleman et Markham, 1991 ; Stone et Rainey, 1991 ; Alku et Vilkman, 1996 ; Alku et coll., 1998 ; Vilkman et coll., 1999).

Les indices morphologiques renvoient, en pratique, aux attributs de forme du signal d'excitation glottique. Des exemples sont le quotient d'ouverture, qui est l'intervalle temporel, pendant lequel la glotte est ouverte, divisé par la durée du cycle ; le quotient « fermant » le temps mis par la glotte à se fermer divisé par la durée du cycle ; le quotient des intervalles ouvrant et fermant ; l'amplitude maximale de la vitesse volumique ainsi que l'amplitude du pic négatif de la vitesse volumique dérivée numériquement (Alku et Vilkman, 1996 ; Vilkman et coll., 2002 ; Bergan et coll., 2004). Un exemple d'un indice morphologique spectral est la « balance » spectrale qui quantifie la richesse harmonique (brillance ou éclat de la voix) (Bergan et coll., 2004). Les indices morphologiques sont le plus souvent utilisés pour décrire les types de phonation.

Les indices morphologiques, supra-segmentaux et de coordination ne sont pas limités aux applications cliniques. Ils ont été largement étudiés par les phonéticiens, linguistes, psychologues et ingénieurs, car ils renseignent sur la production et la perception de la voix et de la parole en général.

Des indices qui ont typiquement, mais pas exclusivement, un rôle clinique décrivent les irrégularités du mouvement des plis vocaux. L'irrégularité renvoie ici à toute déviation de la périodicité mathématique des cycles glottiques. Les causes multiples de ces irrégularités ont été discutées ci-dessus. Ces irrégularités sont décrites temporellement ou spectralement. Les descriptions temporelles reposent sur la mesure des durées et amplitudes des cycles de parole successifs. Les séries temporelles correspondantes sont alors résumées moyennant des indices qui sont souvent définis heuristiquement (Wuyts et coll., 2000). Les descriptions spectrales reposent sur la séparation entre harmoniques et inter-harmoniques. Différents indices combinent différemment harmoniques et inter-harmoniques dans des intervalles spectraux distincts (Hemler et coll., 1997 ; Hsiung et coll., 2002).

Les durées des cycles de parole rapportent, en bonne approximation, les durées des cycles glottiques. En revanche, les perturbations des amplitudes des cycles de parole ont des causes multiples. En effet, la distorsion de modulation du conduit vocal transforme les perturbations des durées des cycles glottiques en perturbations de l'amplitude des cycles de parole. D'autres contributions sont les reports inégaux d'énergie acoustique d'un

cycle à l'autre dans des séquences de cycles de durée inégale, ainsi que le tremblement des articulateurs supraglottiques (Decoster et Debruyne, 1997 ; Wolfe et Martin, 1997 ; Giovanni et coll., 1999 ; Mann et coll., 1999 ; Schoentgen, 2003).

### ***Corrélation***

Souvent, cliniciens ou thérapeutes tentent de corrélérer les valeurs des indices des troubles vocaux avec des données relatives à d'autres niveaux de description. Les données, avec lesquelles une corrélation est recherchée, sont typiquement diagnostiques, aérodynamiques, perceptives, ou basées sur imagerie glottique, avec une préférence pour les données perceptives (Stone et Rainey, 1991 ; Dejonckere et coll., 1996 ; Wolfe et Martin, 1997 ; Alku et coll., 1998 ; Rantala et Vilkman, 1999 ; Habermann et coll., 2000 ; Alku et coll., 2001 ; Hsiung et coll., 2002 ; Yu et coll., 2002 ; Bhuta et coll., 2004).

### ***Catégories de locuteurs***

On observe que beaucoup d'indices acoustiques des troubles vocaux diffèrent pour des locuteurs masculins et féminins, ainsi que pour des locuteurs jeunes et âgés (Decoster et Debruyne, 1997). Les valeurs des indices sont, par conséquent, souvent rapportées séparément pour ces catégories de locuteurs.

### **Problèmes connus**

Les indices acoustiques existants présentent des problèmes connus.

#### ***Traitement du signal***

Les indices vocaux, dont l'usage clinique est répandu, quantifient les irrégularités des cycles vocaux. Des exemples typiques sont les quotients de perturbation de la période et de l'amplitude, la gigue en pourcentage (*jitter in %*), le rapport harmonique à bruit... Très souvent, le traitement du signal repose sur des hypothèses de stationnarité et de périodicité locale qui permettent à des méthodes heuristiques de détecter et d'isoler les cycles de parole ou les harmoniques du spectre. Ces heuristiques peuvent échouer lorsque la voix est sévèrement enrouée. Par conséquent, des erreurs d'insertion ou d'omission de cycles ou d'harmoniques sont fréquentes dans le cas de signaux irréguliers. Ces erreurs biaisent les valeurs des indices mesurés. Ainsi, les indices vocaux qui décrivent des perturbations vocales sont considérés fiables uniquement lorsqu'ils sont extraits de sons voisés soutenus par des locuteurs faiblement ou modérément enroués.

Un deuxième problème est la précision de la mesure des durées des cycles de parole. En effet, les perturbations des durées peuvent être faibles, c'est-à-dire moins de 1 % pour les durées et moins de 10 % pour les amplitudes des cycles. Par conséquent, le traitement du signal est exigeant pour garantir la

précision des mesures qui sont facilement biaisées par le bruit de quantification, le bruit additif...

### **Étiquetage**

L'étiquetage réfère ici à la coutume de donner aux indices vocaux des noms qui font allusion aux symptômes vocaux visés et non aux grandeurs qui ont été mesurées réellement. Par exemple, des indices résumant les dyspériodicités des durées des cycles vocaux sont souvent désignés par le terme de gigue vocale, quoique ces perturbations des durées soient facilement influencées par des facteurs aussi divers que la fréquence phonatoire moyenne, le bruit de turbulence, le tremblement de la fréquence phonatoire, les régimes vibratoires non-modaux et l'intonation.

### **Redondance**

Le nombre d'indices acoustiques qui ont été proposés dans la littérature est large. Des logiciels vendus pour l'évaluation vocale en clinique comprennent typiquement des dizaines d'indices numériques différents qui peuvent être calculés pour chaque son soutenu. Il n'est pas surprenant que des études montrent l'existence d'ensembles d'indices phonatoires corrélés entre eux. Ces ensembles corrélés regroupent typiquement les jeux d'indices qui sont propres à chaque catégorie de symptômes vocaux qui ont été discutés précédemment (Dejonckere et coll., 1996).

### **Interprétation**

Des exemples discutés précédemment au sujet des perturbations des durées et amplitudes des cycles vocaux suggèrent que même les mesures de grandeurs physiques simples sont influencées par de nombreux facteurs. Cette observation, vraie en général, suggère que les indices phonatoires peuvent être difficiles à interpréter car leurs valeurs sont déterminées par des causes multiples interdépendantes.

### **Fragments stationnaires de sons de parole soutenus**

Une critique souvent formulée est que beaucoup d'indices acoustiques sont seulement obtenus avec fiabilité pour des fragments stationnaires de sons soutenus. Les raisons ont été discutées dans la section sur le traitement de signal. Une conséquence est qu'actuellement les effets des pathologies laryngées sur la parole connectée ou naturelle sont mal compris. Les problèmes non résolus ne concernent pas uniquement le traitement de signal, mais aussi le choix des indices phonatoires, le choix du matériau linguistique, ainsi que l'évaluation perceptuelle de fragments de parole connectée qui sont courts ou phonétiquement complexes.

Le tableau 2.I résume quelques facteurs qui différencient les approches de l'évaluation de la voix.

**Tableau 2.1 : Approches de l'évaluation de la voix**

Étiologie	Trouble organique ou dysfonctionnel, trouble moteur, voix de substitution
Transformées (si d'application)	Fourier, Hilbert, transformée en ondelettes
Signaux	Signal glottique, signal de parole
Matériaux	Parole connectée, son soutenu, fragment stationnaire d'un son soutenu
Symptômes	Dyspériodicité et bruit, morphologie du signal, supra-segmental, coordination
Tâches	Parole, chant, chargement vocal, profilage
Performances	Registre, type phonatoire, voisement, prosodie, timbre vocal
Capteurs	Microphone, électroglottographe, photoglottographe, capteur de débit aérien, capteur de pression

**En conclusion,** les pathologies laryngées peuvent provoquer des troubles de la voix dans la mesure où elles affectent les vibrations des plis vocaux. Les troubles de la voix sont évalués perceptivement et/ou acoustiquement, et/ou aérodynamiquement. L'évaluation fonctionnelle acoustique de la fonction laryngée et des troubles de la voix repose sur des indices sonores qui sont extraits du signal microphonique ou d'autres signaux obtenus non-invasivement. Car, la chaîne de causalité qui lie le diagnostic, les motifs vibratoires, les indices sonores et les scores perceptifs est longue et non univoque. L'observation de la vibration est difficile et gêne le locuteur et les troubles de la voix constituent un sujet d'étude semi-autonome dans la mesure où ils forment un obstacle à la communication verbale ou à l'expression esthétique. Les méthodes d'évaluation et les indices sonores sont nombreux car les motivations pour leur développement sont multiples. À l'heure actuelle, l'évaluation perceptive des troubles de la voix et l'évaluation non invasive de la fonction laryngée profitent de nouveaux développements dont l'objectif est d'étendre les conditions sous lesquelles les indices sonores de la fonction laryngée peuvent être observés et enregistrés, ainsi que de diminuer la variabilité résiduelle des scores perceptifs qui est liée aux juges et aux conditions d'évaluation.

## BIBLIOGRAPHIE

ALIPOUR F, SCHERER RC. Vocal fold bulging effects on phonation using a biophysical computer model. *J Voice* 2000, **14** : 470-483

ALKU P, VILKMAN E. A comparison of glottal voice source quantification parameters in breathy, normal and pressed phonation of female and male speakers. *Folia Phoniatr Logop* 1996, **48** : 240-254

ALKU P, VILKMAN E, LAUKKANEN AM. Parameterization of the voice source by combining spectral decay and amplitude features of the glottal flow. *J Speech Lang Hear Res* 1998, **41** : 990-1002

ALKU P, VINTTURI J, VILKMAN E. Evidence of the significance of secondary excitations of the vocal tract for vocal intensity. *Folia Phoniatr Logop* 2001, **53** : 185-197

BÄCKSTROM T, LEHTO L, ALKU P, VILKMAN E. Automatic pre-segmentation of running speech improves the robustness of several acoustic voice measures. *Logoped Phoniatr Vocol* 2003, **28** : 101-108

BEHRMAN A, AGRESTI C, BLUMSTEIN E, LEE N. Microphone and EGG data from dysphonic patients: type 1, 2 and 3 signals. *J Voice* 1998, **12** : 249-260

BERGAN CC, TITZE IR, STORY B. The perception of two vocal qualities in a synthesized vocal utterance: ring and pressed voice. *J Voice* 2004, **18** : 305-317

BERRY D, HERZEL H, TITZE I, STORY B. Bifurcations in excised larynx experiments. *J Voice* 1996, **10** : 129-138

BHUTA T, PATRICK L, GARNETT JD. Perceptual evaluation of voice quality and its correlation with acoustic measurements. *J Voice* 2004, **18** : 299-304

BOONE D, MCFARLANE S, VON BERG S. The voice and voice therapy. Pearson, Boston, 2004

BUNTON K, WEISMER G. Segmental level analysis of laryngeal function in persons with motor speech disorders. *Folia Phoniatr Logop* 2002, **54** : 223-239

COLEMAN RF, MARKHAM IW. Normal variations in habitual pitch. *J Voice* 1991, **2** : 173-177

DECOSTER W, DEBRUYNE F. The ageing voice: changes in fundamental frequency, waveform stability and spectrum. *Acta Otorhinolaryngol Belg* 1997, **51** : 105-112

DEJONCKERE PH, REMACLE M, FRESNEL-ELBAZ E, WOISARD V, CREVIER-BUCHMAN L, MILLET B. Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. *Rev Laryngol Otol Rhinol (bord)* 1996, **117** : 219-224

ELLIOT N, SUNDBERG J, GRAMMING P. What happens during vocal warm-up? *J Voice* 1995, **9** : 37-44

ESSLING J. Voice quality settings of the pharynx. Proc. XIV. Int. Con. Phon. Sc., San Francisco, 1999 : 2449-2452

GIOVANNI A, REVIS J, TRIGLIA JM. Objective aerodynamic and acoustic measurement of voice improvement after phonosurgery. *Laryngoscope* 1999, **109** : 656-660

GIOVANNI A, OUAKNINE M, GARREL R, AYACHE S, ROBERT D. Non-linear model of glottic vibration. Potential clinical implications. *Rev Laryngol Otol Rhinol (bord)* 2002, **123** : 273-277

GUNTER HE. A mechanical model of vocal-fold collision with high spatial and temporal resolution. *J Acoust Soc Am* 2003, **113** : 994-1000

HABERMANN W, JIANG J, LIN E, HANSON DG. Correlation between glottal area and photoglottographic signal in normal subjects. *Acta Otolaryngol* 2000, **120** : 778-782

HEMLER RJ, WIENEKE GH, DEJONCKERE PH. The effect of relative humidity of inhaled air on acoustic parameters of voice in normal subjects. *J Voice* 1997, **11** : 295-300

HEYLEN L, WUYTS FL, MERTENS F, DE BODT M, VAN DE HEYNING PH. Normative voice range profiles of male and female professional voice users. *J Voice* 2002, **16** : 1-7

HILLMAN R, OESTERLE E, FETH L. Characteristics of the glottal turbulent noise source. *J Acoust Soc Am* 1983, **74** : 691-694

HSIUNG MW, PAI L, WANG HW. Correlation between voice handicap index and voice laboratory measurements in dysphonic patients. *Eur Arch Otorhinolaryngol* 2002, **259** : 97-99

ISSHIKI N, KITAJIMA K, KOJIMA H, HARITA Y. Turbulent noise in dysphonia. *Folia Phoniatica* 1978, **30** : 214-224

ISHIZAKA K, FLANAGAN J. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Tech Journal* 1972, **51** : 1233-1268

JIANG JJ, ZHANG Y, STERN J. Modeling of chaotic vibrations in symmetric vocal folds. *J Acoust Soc Am* 2001, **110** : 2120-2128

LADEFOGED P. Vowels and Consonants. Blackwell Publishing, Oxford, UK, 2005

LAUKKANEN AM, TAKALO R, VILKMAN E, NUMMENRANTA J, LIPPONEN T. Simultaneous videofluorographic and dual-channel electroglottographic registration of the vertical laryngeal position in various phonatory tasks. *J Voice* 1999, **13** : 60-71

LUCERO JC. A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset. *J Acoust Soc Am* 1999, **105** : 423-431

MANN EA, MCCLEAN MD, GUREVICH-UVENA J, BARKMEIER J, MCKENZIE-GARNER P, et coll. The effects of excessive vocalization on acoustic and videostroboscopic measures of vocal fold condition. *J Voice* 1999, **13** : 294-302

MORSOMME D, JAMART J, WERY C, GIOVANNI A, REMACLE M. Comparison between the GIRBAS Scale and the Acoustic and Aerodynamic Measures Provided by EVA for the Assessment of Dysphonia following Unilateral Vocal Fold Paralysis. *Folia Phoniatr Logop* 2001, **53** : 317-325

OHLSSON AC, BRINK O, LOFQVIST A. A voice accumulation-validation and application. *J Speech Hear Res* 1989, **32** : 451-457

PINTO NB, TITZE IR. Unification of perturbation measures in speech signals. *J Acoust Soc Am* 1990, **87** : 1278-1289

RANTALA L, VILKMAN E. Relationship between subjective voice complaints and acoustic parameters in female teachers' voices. *J Voice* 1999, **13** : 484-495

REVIS J, BARBERIS S, GIOVANNI A. Definition of a new temporal voice onset measurement. *Rev Laryngol Otol Rhinol (bord)* 2000, **121** : 291-296

ROTHMAN HB, BROWN WS JR, SAPIENZA CM, MORRIS RJ. Acoustic analyses of trained singers perceptually identified from speaking samples. *J Voice* 2001, **15** : 25-35

SCHOENTGEN J, BENSALD M, BUCELLA F. Multivariate statistical analysis of flat vowel spectra with a view to characterizing dysphonic voices. *J Speech Lang Hear Res* 2000, **43** : 1493-1508

- SCHOENTGEN J. Modulation frequency and modulation level owing to vocal micro-tremor. *J Acoust Soc Am* 2002, **112** : 690-700
- SCHOENTGEN J. Spectral models of additive and modulation noise in speech and phonatory excitation signals. *J Acoust Soc Am* 2003, **113** : 553-562
- SHEARER WM. Illustrated Speech Anatomy. Charles C. Thomas, Publisher, Springfield, Illinois, 1963
- SPENCER M, TITZE I. An investigation of a modal-falsetto register transition hypothesis using helox gas. *J Voice* 2001, **15** : 15-24
- STEMPLE JC. Voice research: so what? A clearer view of voice production, 25 years of progress; the speaking voice. *J Voice* 1993, **7** : 293-300
- STEVENS K. Acoustic Phonetics. Springer, Cambridge, MA, 1998
- STONE RE, RAINEY CL. Intra and intersubject variability in acoustic measures of normal voice. *J Voice* 1991, **5** : 189-196
- SUNDBERG J. The science of the singing voice. Northern Illinois University Press, Dekalp, Illinois, 1987
- TILL J, YORKSTON K, BEUKELMAN D. Motor speech disorders, Advances in Assessment and Treatment. Paul Brookes Publ, Baltimore, 1994
- TITZE I. Principles of Voice Production. Prentice Hall, Englewood Cliffs, New Jersey, 1994
- TITZE I, SCHMIDT X. Phonation threshold pressure in a physical model of the vocal fold mucosa, *SPR-NCVS* 1994, **6** : 13-18
- VILKMAN E, LAURI ER, ALKU P, SALA E, SIHVO M. Loading changes in time-based parameters of glottal flow waveforms in different ergonomic conditions. *Folia Phoniatr Logop* 1997, **49** : 247-263
- VILKMAN E, LAURI ER, ALKU P, SALA E, SIHVO M. Effects of prolonged oral reading on F0, SPL, subglottal pressure and amplitude characteristics of glottal flow waveforms. *J Voice* 1999, **13** : 303-312
- VILKMAN E, ALKU P, VINTTURI J. Dynamic extremes of voice in the light of time domain parameters extracted from the amplitude features of glottal flow and its derivative. *Folia Phoniatr Logop* 2002, **54** : 144-157
- WOLFE V, MARTIN D. Acoustic correlates of dysphonia: type and severity. *J Commun Disord* 1997, **30** : 403-415
- WUYTS FL, DE BODT MS, MOLENBERGHS G, REMACLE M, HEYLEN L, et coll. The dysphonia severity index: an objective measure of vocal quality based on a multiparameter approach. *J Speech Lang Hear Res* 2000, **43** : 796-809
- YU P, OUAKNINE M, REVIS J, GIOVANNI A. Objective voice analysis for dysphonic patients: a multiparametric protocol including acoustic and aerodynamic measurements. *J Voice* 2001, **15** : 529-542
- YU P, REVIS J, WUYTS FL, ZANARET M, GIOVANNI A. Correlation of instrumental voice evaluation with perceptual voice analysis using a modified visual analog scale. *Folia Phoniatr Logop* 2002, **54** : 271-281