

Pourquoi rapprocher langue et musique ?

« Toute oreille, même non musicienne, sait discerner le rythme de la phrase, les inflexions de son intonation, les accentuations et les fluctuations de son intensité, comme les formes et les variantes de son timbre puisque c'est leur assemblage qui fait sens. Autrement dit, dans la parole, tout est déjà composé [...] On le sait, un flux parlé génère une mélodie, celle de la langue à travers son locuteur. Aujourd'hui, les outils d'analyse du son montrent qu'elle contient aussi une harmonie, une harmonie naturelle en sorte », Claudy Malherbe (disponible sur le site : <http://sol.ircam.fr/instruments/musicweb.fr>).

Table du premier chapitre

1. Éléments pour une analogie entre langue et musique

- 1.1. Un parallèle implicite posé entre prosodie et musique
- 1.2. Similitudes et divergences de forme
 - 1.2.1. Rythme musical et rythme verbal
 - 1.2.1.1. Définition commune
 - 1.2.1.2. Musique et métricité
 - 1.2.1.3. Parole et périodicité sommaire
 - 1.2.2. Mélodie de la musique et mélodie de la parole
 - 1.2.2.1. Successivité et intervalles
 - 1.2.2.2. Le fait musical au travers du filtre de la transcription
 - 1.2.2.3. Des précautions à prendre
 - 1.2.2.4. Tension tonale dans la musique et dans la langue
 - 1.2.2.5. Différenciation mélodie langue / musique
 - 1.2.3. A mi chemin entre musique et langue : les clichés mélodiques
 - 1.2.4. Degré de mélodicité et de rythmicité, la clé de la musicalité ?
- 1.3. Cas de fusion avec la musique : les langues « musicales »
 - 1.3.1. Langues sifflées
 - 1.3.2. Langues tambourinées
 - 1.3.3. Langue artificielle musicale
- 1.4. L'origine de la langue dans la musique ?

2. Exploitation des similitudes de forme

- 2.1. Lorsque la musique devient langue
 - 2.1.1. Transcription musicale de la prosodie
 - 2.1.1.1. Potentialités de la transcription musicale
 - 2.1.1.2. Historique des applications de la transcription musicale en prosodie
 - 2.1.1.3. Transfert de la terminologie musicale vers la prosodie
 - 2.1.2. D'autres systèmes de transcription : leur validité par rapport à la transcription musicale
- 2.2. Lorsque la langue devient musique : Evolution du traitement de la similitude entre la langue et la musique au cours de l'histoire de la musique occidentale
 - 2.2.1. Rapports texte et mélodie : information vs musicalité
 - 2.2.1.1. Le chant grégorien : une musique du texte
 - 2.2.1.2. Le récitatif dans l'opéra :
 - 2.2.1.2.1. A la frontière de la parole et de la musique
 - 2.2.1.2.2. Trois récitatifs, trois langues, trois musiques
 - 2.2.1.2.1.1. Anglais : « *Didon et Énée* » de Purcell
 - 2.2.1.2.1.2. Allemand : Les « *Passions de St Matthieu* » de Bach
 - 2.2.1.2.1.3. Italien : « *Don Giovanni* » de Mozart

- 2.2.1.3. La musicalité des langues dans la mise en musique :
 - 2.2.1.3.1. Langues musicales versus langues non musicales ?
 - 2.2.1.3.2. Un obstacle à la traduction des œuvres musicales ?
- 2.2.2. La crise de la mélodie ou le retour à l'intonation
 - 2.2.2.1. Debussy et la musicalité de la langue française
 - 2.2.2.2. Le « Sprechgesang »
- 2.2.3. Musicalisation de la parole dans la création contemporaine
 - 2.2.3.1. La langue traitée comme matériau : la voix chez Boulez et Berio
 - 2.2.3.2. Le musicien « phonéticien »
 - 2.2.3.3. Typologie des effets « langue » dans la musique

3. Similitudes de formes étendues à des similitudes « structurales »

- 3.1. Universaux
- 3.2. Similitudes structurales
- 3.3. Peut-on parler de grammaire et de langage à propos de la musique ?
- 3.4. Capacité cognitive commune

Bibliographie du chapitre I

Si notre intuition nous suggère une situation de proximité maximale entre les éléments prosodiques et la musique, il convient de transformer cette intuition en une description précise des similitudes de forme qu'elles entretiennent, qui viendra justifier l'analogie. Mieux connaître ce qui rapproche le rythme verbal du rythme musical et la mélodie de la langue de celle de la musique, c'est également mieux identifier la frontière ténue qui les sépare. Cette frontière devient très floue dans les clichés mélodiques et elle s'abolit complètement dans les langues « musicales », qui n'utilisent plus que le niveau musical comme vecteur du niveau linguistique. Des linguistes et des philosophes ont d'ailleurs vu l'origine de la langue dans la musique, certains posant l'hypothèse d'une origine commune. Face au manque de système de notation pour les éléments musicaux de la langue, c'est donc tout naturellement que les phonéticiens ont fait appel à la transcription musicale et l'ont adaptée à leur objet. A l'inverse, les similitudes avec la langue ont toujours inspiré les compositeurs et les rapports entre le « texte » et la « musique » constituent même la trame de fond de l'histoire de la musique vocale occidentale. L'étude détaillée de ces rapports nous permettra de compléter et d'enrichir notre approche de phonéticienne grâce aux réflexions très subtiles menées depuis des siècles par les spécialistes de l'autre « versant ». Enfin, il n'est guère étonnant que ces similitudes de formes aient à leur tour attiré les théoriciens qui ont cherché à les étendre à des similitudes structurales.

1. Éléments pour une analogie entre langue et musique

1.1. Un parallèle implicite posé entre prosodie et musique

L'existence de liens étroits entre musique et prosodie a poussé certains chercheurs à parler d'éléments musicaux pour désigner le rythme et l'intonation, notamment Faure (1962 : 7) dans sa recherche sur le rôle des éléments musicaux dans la prononciation anglaise :

« Au cours de la présente étude, nous entendrons par éléments musicaux de la prononciation anglaise tout ce qui, dans cette langue, se combine à l'articulation des sons pour permettre au discours d'assurer la communication d'un certain contenu de pensée ou d'émotions. Ces éléments sont essentiellement constitués par les variations de hauteur qui affectent, de diverses façons, les syllabes successives d'un énoncé et qui, s'accompagnant presque toujours de variations concomitantes d'intensité et de durée, en déterminent à la fois le rythme et la mélodie ».

Or, parler d'éléments musicaux à propos de la langue, c'est poser implicitement un parallèle entre musique et langue. Et selon Menezes (1993 : 81) :

« du point de vue de la réception linguistique, c'est indubitablement à travers l'intonation prosodique que l'on s'aperçoit de façon immédiate de la contiguïté innée entre langage et musique ».

En se penchant sur l'étymologie de ces termes, il est d'ailleurs frappant de constater qu'ils ont une origine musicale. En effet, le terme « prosodie » provient de l'ancien grec « ôdê », qui veut dire chant ou plus précisément chant accompagné d'une musique instrumentale. Il a ensuite été utilisé pour désigner la « science de la versification » et la métrique, qui gouverne la voix humaine lorsqu'elle est en train de lire de la poésie. Quant au terme « intonation », il a lui aussi une origine musicale : il vient du latin « intonare », qui signifie tonner, faire retentir¹. Rossi (1999) explique que :

*« par une fausse étymologie avec tonus (ton), intonation est à l'origine un terme musical utilisé pour désigner l'action d'entonner un air. Ce n'est guère avant le XIXe siècle qu'on l'utilise également pour désigner les tons de la parole. A la fin du XVIIIe, son usage linguistique n'est pas encore consacré, puisque Walker, en 1787, dans son ouvrage, *The Melody of Speaking Delineated*, consacré à l'intonation de l'anglais, ne l'utilise pas encore. Au XIXe, le terme intonation désigne, chez Chateaubriand, « les tons de la voix » ; il est donc synonyme de musicalité, mélodie ; déjà le glissement de sens est manifeste par rapport à la signification originelle».*

Pour Scripture (1902 : 472), la description de la mélodie de la parole devrait prendre en compte le traitement de la mélodie dans le chant, et spécialement les chants primitifs et les chants spontanés des enfants².

¹ Intonation (1372) : « Golein, du latin intonare, tonner, faire retentir, rattaché par fausse étymologie à tonus, ton ». Dictionnaire Étymologique, Larousse, 1964.

² Scripture (1902 : 472) : « a discussion of the melody of speech should, perhaps, include a treatment of melody in song, particularly the primitive song of uncultured peoples and the spontaneous song of children ».

S'intéressant justement aux premières productions vocales de l'enfant, qu'elles soient langagières ou musicales, Papousek et Papousek (1981 : 166) soulignent cette interdépendance entre langage et musique :

« The evolution and interdependence of language and music are phenomena common to all human cultures, and their universality need not be documented ».

Et pour mieux asseoir ce parallèle, ils parlent également d'éléments musicaux pour désigner les éléments prosodiques (1981 : 164), exactement comme le faisait Faure :

« Musical elements – pitch, tone, interval, melody, loudness, accent, tempo and rhythm – or their acoustic correlates – frequency, harmonics, amplitude and temporal patterns – are matters of a primary interest to musicologists but have not attracted much attention of linguists or phoneticians studying language acquisition³ ».

L'utilisation de l'expression « éléments musicaux » permet de concilier l'approche linguistique et l'approche « musicologique », cette double approche étant nécessaire à la description linguistique, tout au moins au début de la vie :

« Dealing in details with the infant's earliest perception and production of vocal elements which represent musical element to a musicologist, suprasegmental features to a phonetician, and prosodic or paralinguistic elements to a linguist, we are going to speak of "musical" rather than "prosodic" elements. Thus we want to facilitate our discussion of both linguistic and musicological aspects. For the same purpose our procedures include musical transcriptions as well » (Papousek et Papousek, 1981 : 164).

1.2. Similitudes et divergences de forme

Si le parallèle posé entre la langue et la musique est né des similitudes de forme qu'elles entretiennent, encore faut-il décrire ces similitudes. Mais la plupart des auteurs ne se donnent pas la peine de le faire ou alors se contentent de quelques arguments donnés au compte-goutte. Il s'agit donc de justifier le fait qu'on puisse parler de musique à propos de la langue, et plus précisément, de déterminer ce qui est commun au rythme musical et au rythme verbal et ce qui est commun à la mélodie de la musique et à la mélodie de la parole.

³ Après 1980, les phonéticiens ont comblé leur retard et nous verrons dans le chapitre II que de nombreux travaux ont été menés sur la mise en place des éléments musicaux dans l'acquisition du langage.

1.2.1. Rythme musical et rythme verbal

1.2.1.1. Définition commune

Qu'est-ce que le rythme ? A la base de la sensation de rythme, il faut une séquence d'au moins deux éléments qui se distinguent sur le plan perceptif. Comment certains éléments se distinguent-ils des autres ? Parce qu'ils sont marqués par une proéminence, par un accent (l'accent est défini comme un « *fait local de proéminence* » par Di Cristo, 1999 : 185). Comment se manifeste cette proéminence ? Par un agencement spécifique des paramètres acoustiques que sont la fréquence, la durée et l'amplitude, qui va engendrer une mise en relief au niveau perceptif (hauteur, durée et intensité) et au niveau syntagmatique. C'est donc l'alternance des éléments saillants et des éléments non saillants qui va donner naissance au rythme. Selon la définition du Littré, le rythme dans la langue se définit par une **succession** de syllabes **accentuées** (sons forts) et de syllabes non **accentuées** revenant à certains intervalles. Dans ce sens, les syllabes constituent donc des unités rythmiques.

« Le rythme est fondé sur l'organisation temporelle des syllabes métriquement fortes et des syllabes métriquement faibles », (Di Cristo, 1999 : 186).

Appliqué à la musique, il désigne une **succession** de **temps forts** et de **temps faibles**. La définition est donc similaire pour la langue et pour la musique. Si l'on revient à l'étymologie du mot, *ruthmos* (du grec *rein*, qui signifie « couler » au sens d'un ruisseau qui coule) signifie chez les auteurs grecs :

« forme distinctive, figure proportionnée, disposition », ou encore « arrangement caractéristique des parties dans un tout », « manière particulière de fluer »,

et, chez Platon, pour la première fois,

« forme du mouvement que le corps humain accomplit dans la danse » (Benveniste, 1951).

Sauvanet (2002) pose donc la définition globale suivante :

*« tout phénomène, **perçu ou agi**, auquel on peut attribuer au moins deux des qualités suivantes: structure, périodicité, mouvement» et la définition restreinte : « tout phénomène, perçu ou agi, auquel on peut attribuer chacune de ces trois qualités ».*

Cette définition s'applique à la fois à la production et à la perception de la musique et de la langue. Un rythme fort combine les trois : la musique et la langue sont des **rythmes forts**. Elles combinent la structure (la forme, avec l'agencement du tout et des parties), la périodicité (le retour régulier de temps forts ou des syllabes accentuées) et le mouvement (exécution musicale et parole). Mais, dans certaines pièces de musique contemporaine, on trouve une structure, un mouvement, mais pas forcément de périodicité.

Il existe une tradition de description du rythme musical qui s'enracine dans ces similitudes entre rythme verbal et rythme musical. En effet, elle utilise un système basé sur les pieds de la poésie pour décrire le rythme musical. Cette tradition remonte au temps d'Aristoxène et

« il est possible d'en suivre la trace depuis les théoriciens de la Renaissance et du XVIIIe, jusqu'à Riemann, le théoricien de la musique du XIXe » (Clarke, 1989 : 26).

Plus récemment, elle fonde la description que font Cooper et Meyer (1960) de la structure rythmique de la musique : ils ont en effet constitué un répertoire de cinq types rythmiques inspirés de la prosodie grecque (le trochée, le dactyle, l'iambe, l'anapeste et l'amphibraque), définis selon la répartition des éléments accentués et non accentués (tableau n°1).






Trochée	Dactyle	Iambe	Anapeste	Amphibraque
				

Tableau n°1 : figures rythmiques correspondant aux cinq types rythmiques définis par Cooper et Meyer (1960).

1.2.1.2. Musique et métricité

Malgré ces similitudes, le cadre rythmique de la musique est pourtant différent de celui de la langue. Selon Jones et Yee (1994 : 108), la musique occidentale

« semble nous inviter à frapper le rythme avec le pied et à le suivre, ce qui suggère une apparente métricité ».

Sur la partition musicale, les structures sont effectivement organisées de façon mesurée : des points d'appui périodiques délimitent les mesures, la pulsation se définissant comme le retour régulier de ces points d'appui. Chacun d'entre eux est associé à un temps en musique.

Afin de rendre le rythme lisible, la notation musicale répartit ces temps au sein d'unités temporelles de taille égale, les mesures qui sont délimitées par des barres de mesure. On trouve ainsi des mesures comportant 2, 3, 4, 5 temps, etc... Selon Clarke (1989 : 27) :

« le rythme s'organise autour d'un cadre d'unités isochrones répétées (la pulsation), de telle sorte que de grandes périodes temporelles sont reliées l'une à l'autre par des rapports de nombres entiers. On obtient ainsi un cadre stable qui, entre autres choses, facilite la coordination des différents interprètes dans les ensembles ».

Lorsque ces unités sont isochrones et reliées entre elles par des rapports temporels simples, les auditeurs sont capables de se synchroniser avec le rythme et d'en extraire facilement la pulsation en la frappant (Michon, 1967 ; Wing et Kristofferson, 1973a, b), et ce, même après la fin de la séquence. Povel (1981) a montré en les chronométrant, que les battements coordonnés avec le plus de précision étaient ceux qui reproduisaient un rapport temporel simple et notamment, le plus simple d'entre eux, le rapport 2:1.

Les auditeurs sont donc capables d'engendrer et d'appliquer des relations métriques pour lier des groupes et des sons. Mais ils vont plus loin qu'extraire la seule pulsation. Nous avons vu que le rythme musical se définissait comme une succession de temps forts et de temps faibles. Quel que soit leur « poids » respectif, ces temps s'écoulaient régulièrement. Dans une mesure, le premier temps est toujours fort, le deuxième et le troisième sont faibles, sauf dans la mesure à quatre temps où le troisième est fort. Chaque temps se divise lui-même en une partie initiale forte et une partie finale faible. Après avoir extrait la pulsation, les auditeurs, sensibles à cette hiérarchie rythmique, sont capables de repérer le ou les temps forts au sein de la mesure.

« Dans le cas d'un rythme de valse par exemple, l'auditeur repérera tout d'abord un premier niveau de battement régulier au niveau de la noire, puis il interprétera un de ces battements sur trois comme étant le plus important » (Bigand, 1994 : 273).

Pour un même morceau de musique, l'auditeur peut extraire une pulsation différente, dont les battements seront plus rapprochés (tempo rapide) ou au contraire plus éloignés (tempo lent). Ces différences de pulsation reflètent en réalité des niveaux hiérarchiques différents. Lerdahl et Jackendoff (1983) ont montré que la musique répondait à une organisation hiérarchique d'unités de durées variables : au plus bas niveau, les unités sont constituées d'un seul élément.

Ces unités se combinent ensuite pour créer d'autres unités plus longues, à des niveaux plus élevés. Plus l'auditeur sera expérimenté (acculturation avec le système musical classique et niveau de formation musicale), plus il aura accès à un grand nombre de niveaux. Il pourra ainsi organiser les événements sur de plus grands groupes temporels et « choisira » donc spontanément une pulsation plus lente, reflétant un niveau hiérarchique élevé (Drake et al., 2000a). La figure n°1 extraite de Drake et al. (2000b : 253) nous donne un exemple des différents niveaux de hiérarchie qui affectent un même extrait musical. Un musicien aura accès aux niveaux les plus élevés, c'est-à-dire qu'il sera capable de battre la pulsation tous les deux temps (n+1) ou tous les quatre temps (n+2).

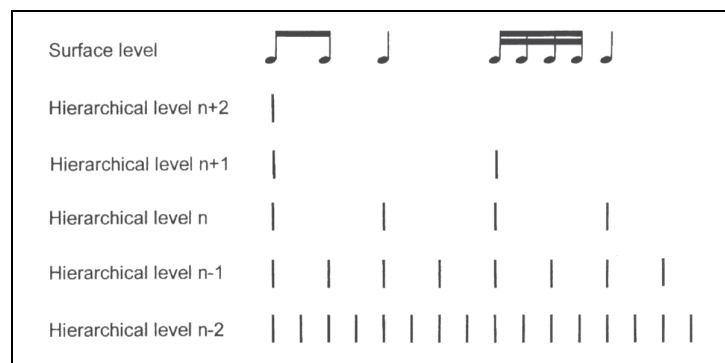


Figure n°1 (emprunté à Drake et al., 2000b : 253).

Il reste cependant un paradoxe au sein d'une organisation temporelle, qui paraît aussi réglée que la mécanique d'un horloge suisse. Alors que l'organisation mesurée de la musique et son organisation temporelle hiérarchique apparaissent sans ambiguïté sur la partition musicale, il n'en va pas de même au niveau de l'interprétation musicale. L'interprète ne peut jouer avec le paramètre de durée, sans risquer de violer le cadre rythmique. Cependant, il semble que l'auditeur soit sensible à des différences minimales de durée entre les notes. Dans une séquence composée de notes de durées semblables, des différences de durée aussi petites que 10 ms peuvent être détectées (Van Noorden, 1975). Les musiciens semblent profiter de cette forte sensibilité à la perturbation des durées chez l'auditeur pour communiquer l'accentuation. S'ils prennent la liberté de marquer l'accentuation par un ajout de durée, cela produira un ralentissement du tempo.

Le musicien virtuose joue avec le cadre rythmique qui lui est imposé, jusqu'aux limites de ce cadre, mais sans jamais le violer ; la musicalité est à ce prix et il semble bien que l'oreille n'apprécie pas les rythmes trop réguliers, qui lui paraissent non naturels :

« Les interprètes contrôlent intentionnellement leurs réactions pour créer des patterns sonores musicaux qui entravent systématiquement les attentes des auditeurs concernant les relations temporelles hiérarchiques ; ils donnent de ce fait une interprétation de la partition » (Jones et Yee, 1994 : 109).

La régularité qui apparaît sur la partition n'existe donc dans la réalité de l'interprétation. Pourtant, nous avons vu qu'elle constituait une réalité psychologique pour l'auditeur, qui se montre capable d'extraire une pulsation et ce, malgré une forte sensibilité aux petites différences de durée. Le terme de « métrique » désigne ce processus d'abstraction :

« La métrique est donc une abstraction, une idéalisation, qui implique l'existence de processus cognitifs complexes conduisant l'auditeur à interpréter comme régulières des successions d'événements sonores qui ne le sont objectivement pas (si ce n'est dans le cas des exécutions musicales par ordinateur) » (Bigand, 1994 : 273).

Ce processus joue un grand rôle, puisqu'il crée des attentes perceptives chez l'auditeur : celui-ci attend le retour des temps forts. Certaines figures rythmiques jouent avec ces attentes : c'est notamment le cas du **contretemps** (temps fort ou partie forte du temps occupé par un silence et partie faible, par une note ; ceci viole l'attente que la note tombe sur la partie forte) et de la **syncope** (note qui commence sur un temps faible et se poursuit grâce à une **liaison rythmique**, sur un temps fort ; il en résulte une tension rythmique qui ne peut se résoudre que sur le temps fort).

Comment l'auditeur réussit-il à restituer la pulsation d'un morceau de musique et à accéder à différents niveaux qui reflètent une organisation temporelle hiérarchique ? Une telle organisation implique en effet une organisation régulière. Jones et Yee (1994 : 109) tentent de résoudre cette énigme en pointant le rôle de l'attention dans la perception des patterns rythmiques. Les auditeurs expérimentés apprennent à extraire certaines propriétés rythmiques de la musique, qu'ils projettent ensuite sur les pièces qu'ils entendent. Une étude de Palmer et Krumhansl (1990) citée par Jones et Yee (1994 : 111) révèle ce pouvoir « organisateur » de l'attention.

Les auteurs font entendre à des auditeurs musiciens et non musiciens une séquence de 9 battements isochrones. Un battement sonde d'une hauteur différente a été intercalé entre les cinq derniers battements. La tâche des auditeurs consiste à évaluer le degré d'intégration de ce battement au sein de la séquence rythmique. Si les auditeurs sont informés qu'ils vont entendre un rythme binaire (2 ou 4 battements), ils jugent que le battement intervenant près du début des groupes métriques binaires est très intégré au rythme de la phrase. Mais, s'ils sont informés qu'ils vont entendre un rythme ternaire (3 ou 6 battements), ils jugent que c'est le battement qui intervient près du début des groupes rythmiques ternaires qui est le plus intégré. Il semble donc que l'attention soit fondamentale, puisqu'elle permet d'orienter la perception du rythme, de prédisposer à faire certains groupements, reposant sur des hiérarchies temporelles. Cet effet de « prédisposition attentionnelle » semble d'ailleurs plus prononcé chez les musiciens, ce qui soulignerait le rôle de l'expérience musicale dans la formation de l'attention. Par ailleurs, les musiciens se montrent capables de compléter des airs folkloriques fondés sur des rythmiques simples (Jones et Boltz, 1989) et cette tâche est beaucoup plus précise lorsque ces airs sont structurés hiérarchiquement.

1.2.1.3. Parole et périodicité sommaire

Même si l'interprétation musicale ne se montre pas parfaitement régulière, elle est donc perçue par l'auditeur comme métrique. Il n'en va pas de même pour l'organisation temporelle de la langue. Il n'est qu'à essayer de battre la pulsation en même temps qu'un auditeur parle pour se rendre compte que la tâche est beaucoup plus difficile que pour la musique (Drake, 2002). Le rythme de la langue paraît plus complexe, mais cette complexité dépend également du type de discours. Le rythme peut être en effet très scandé et régulier, comme dans le cas de la récitation (prières), des sermons, de la poésie, du théâtre, des discours politiques... Mais nous nous arrêterons au rythme de la parole dans la conversation courante. Ce rythme paraît non régulier et il semble fondé sur une sorte de périodicité sommaire :

« L'organisation temporelle de la langue parlée n'est toujours pas bien comprise, mais il semble que, bien qu'elle se fonde sur une sorte de périodicité sommaire (ou « quasi-périodicité »-near-miss-periodicity-, selon Schaffer, 1982) et impose de fortes contraintes à certains aspects de la structure temporelle, elle n'utilise pas le même cadre métrique que celui qu'on trouve en musique » (Clarke, 1989 : 27-28).

Selon Fraisse (1974), le retour périodique d'une différence acoustique tend à être interprété comme un temps fort de la structure métrique. Mais quel est cet élément qui revient « quasi » périodiquement ? Il semble qu'il ne soit pas le même selon les langues et il est possible de classer les langues en fonction du type d'élément accentué qui assure cette périodicité : ainsi, il existerait selon Pike (1945) des langues à rythmicité syllabique (« syllable timed ») et des langues à rythmicité accentuelle (« stress timed »). Cependant, on pense aujourd'hui que les différentes langues, au lieu d'appartenir à l'une de ces catégories de manière nettement délimitée, se répartissent le long d'un continuum allant de la rythmicité syllabique à la rythmicité accentuelle (Dauer, 1987). Ainsi, on peut dire que l'anglais a une tendance à la rythmicité accentuelle car cette langue est rythmée par le retour régulier de l'accent tonique (le fameux « *drum beat* », voir chapitre V, 2.1.1.3.). La durée des groupes temporels entre les accents toniques est donc « invariable ». En fait, il serait plus juste de dire qu'elle est perçue comme invariable par l'auditeur, car on note des différences de durée relativement importantes entre les groupes au niveau acoustique. Les auditeurs attendent donc une régularité rythmique et il semble que les locuteurs jouent de cette attente lorsqu'ils perturbent cette régularité, en insérant des pauses ou des accents là où on ne les attend pas par exemple.

Cette attitude se rapproche beaucoup de celle du musicien lorsque celui-ci accentue certains éléments en augmentant légèrement leur durée. Musique et langue se rejoignent finalement dans le fait qu'elles tirent toutes deux de l'expressivité lorsqu'elles jouent et altèrent la métricité. Cependant, la marge de manœuvre semble plus grande pour la parole, car le locuteur n'a pas besoin d'adhérer à un cadre temporel strict. Il peut donc jouer plus librement avec les différents paramètres pour la réalisation de l'accent, dans la mesure où il respecte certaines règles que lui impose sa langue maternelle (durée pour le français ; hauteur, intensité et durée pour l'anglais). Dans la musique classique, cette marge de liberté est beaucoup plus réduite que dans la langue car l'emploi de l'accent est régi par de fortes contraintes, notamment au niveau de la durée. Les chercheurs ne savent pas encore exactement comment le rythme de la parole est perçu, ni quelles unités marquent les niveaux inférieurs et supérieurs de la structure temporelle. Selon Jones et Yee (1994 : 118), les hiérarchies temporelles sont peut-être plus complexes pour la parole que celles de la musique, et donc plus difficiles à faire apparaître.

1.2.2. Mélodie de la musique et mélodie de la parole

1.2.2.1. Successivité et intervalles

Dans l'Encyclopedia Universalis, la mélodie est définie comme :

« une succession de sons ayant entre eux des rapports d'intervalles et de durée. Telle est la définition la plus large et la plus indéterminée qui puisse en être donnée » (Barraud, 2002).

Cette **succession de sons** correspond au niveau acoustique à l'évolution de la fréquence fondamentale au cours du temps et au niveau de la perception, à l'évolution de la hauteur au cours du temps.

« Comme la musique, le langage est sonore et se déroule dans le temps. Il est intéressant de comparer les emplois, structures et perceptions qui divergent à partir de cette base commune » (Schaeffer, 1966 : 284).

Selon Menezes (1993 : 82), le jeu des oppositions est syntagmatique à la fois pour les traits prosodiques et pour les « traits musicaux » et la séquence temporelle est la condition de leur existence :

« la séquence ne jouerait-elle pas à vrai dire le rôle le plus important aussi bien pour la musique que pour le langage ? ».

Langue et musique se caractérisent donc par leur **dimension temporelle** et leur **caractère séquentiel**. Cette similitude entre la structure des trames temporelles en musique et la structure prosodique du langage est également soulignée par Lerdahl et Jackendoff (1983). Dans la musique et dans la langue, les sons qui composent la ligne mélodique entretiennent entre eux des rapports d'**intervalles**, c'est-à-dire une distance entre leurs hauteurs respectives. Cependant, c'est dans la transition entre les sons que réside « la » différence toujours invoquée entre la langue et la musique : en musique, le passage d'une note à une autre se fait de manière discontinue, alors que dans la langue, il se fait de manière progressive et continue, de telle manière que les hauteurs ne peuvent être clairement délimitées et isolées par l'oreille.

D'après Scripture (1902 : 472), **Aristoxène**, le premier musicologue européen avait déjà signalé cette différence :

« It was pointed by Aristoxenus that the difference between the tones in song and those in speech lies in the fact that the voice in singing proceeds by jumps from one note to another, while in speech it continually slides up and down ».

Passy (1890 : 68) cite deux exemples de transcription musicale, le premier (à gauche de la figure n°2) d'un extrait musical et le second (à droite de la figure n°2), d'un extrait de parole. Ce faisant, il tente de démontrer que la transcription musicale se montre inapte à transcrire la mélodie de la parole, car elle ne permet pas de représenter les glissements qui la caractérisent.



Figure n°2 : transcription musicale de l'énoncé « Viens tu ? » dans une version musicale et dans la parole (emprunté à Léon, 1992 : 124).

1.2.2.2. Le fait musical au travers du filtre de la transcription

Cette différence pousse Crystal à affirmer que l'analogie entre langue et musique n'est pas vraiment fondée (1987 : 173). Mais, avant d'en arriver à une telle conclusion, il est nécessaire de prendre certaines précautions et en premier lieu, en ce qui concerne le système de notation musical. La plupart des auteurs oublient que ce système n'est pas la musique, mais bel et bien une **représentation** de la musique. C'est comme si on décrivait la langue parlée en se référant à l'écrit. Et comme toute représentation, elle obéit à un ensemble de critères qui ne sont pas objectifs. En effet, la notation musicale telle que nous la connaissons (voir figure n°17) est profondément ancrée dans une tradition musicale spécifique, la musique occidentale classique. Celle-ci a été fortement marquée par la pratique de l'écriture, à tel point que composition et écriture ont fini par se confondre : le compositeur dispose en effet d'un système d'écriture tellement puissant qu'il lui permet d'agencer des partitions d'une telle complexité qu'aucun musicien ne saurait les interpréter. L'étymologie du mot composition renvoie d'ailleurs à une idée d'arrangement graphique plutôt qu'à un arrangement des sons en fonction du temps. Il est donc dangereux de penser que le système de codification actuel, qui n'a guère évolué depuis le milieu du XVIIIe siècle et qui tend aujourd'hui à s'universaliser, soit le mieux approprié au langage musical, d'autant plus que ce système tend à privilégier la fixation de la hauteur et de la durée.

Au travers du filtre de la notation classique, on en arrive donc à la conclusion, trop simpliste à notre goût, que la musique bondit, alors que la langue, tel un skieur élégant, glisse d'une note à l'autre.

Et bien souvent, cette constatation suffit à enterrer définitivement toute discussion sur les similitudes de forme entre langue et musique. Les choses sont souvent bien moins simples qu'elles y paraissent au premier abord et il suffit pour s'en convaincre de nous projeter, pour un bref voyage, dans l'histoire de la notation musicale occidentale. Et de découvrir qu'à son origine était... la langue.

Ainsi, Massin et Massin (1983 : 126) nous apprennent qu'au II^e siècle de notre ère, Aristophane de Byzance avait mis au point un système de notation pour la langue grecque, où l'accent aigu représentait l'élévation de la voix, l'accent circonflexe, une faible élévation, le point, la chute puis l'arrêt de la voix à la fin d'une phrase. D'ailleurs, d'après son étymologie, le terme même d'accent renvoie à l'origine de la constitution mélodique⁴. Ainsi, les symboles graphiques que nous utilisons aujourd'hui pour transcrire « é, è et ê » étaient utilisés à l'origine pour représenter les modulations de la voix. Au VII^e siècle est apparue une notation utilisée pour transcrire la prosodie des textes de la Bible, et notamment celle du grec et du latin : la notation neumatique. Lors de la lecture des Écritures, la mélodie était plus proche de la déclamation que du chant et les neumes représentaient l'évolution de la ligne mélodique sous forme de « gestes » stylisés graphiquement et probablement inspirés des gestes du chef de chœur dans la cantillation byzantine.

« Cette mise en relief « spatialisée » des inflexions vocales dans des formes de notation neumatique est d'ailleurs présente dans d'autres traditions que la nôtre (cf. la psalmodie bouddhique, le nô japonais, etc.) » (Massin et Massin, 1983 : 126).

⁴ « Du latin *accentus*, élévation de la voix sur une syllabe, puis son d'un instrument » (Dictionnaire étymologique Larousse).

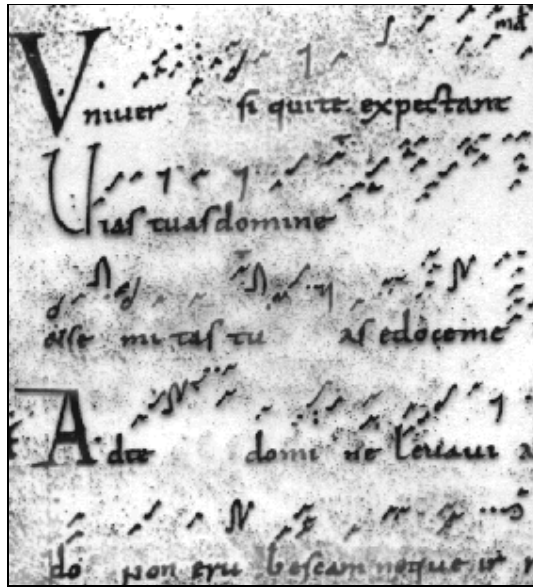


Figure n°3 : Fragment de neumes du manuscrit de Laon 239 - Metz vers 930 (emprunté à Schuyesmans, 2002⁵).

Les neumes sont inscrits au dessus du texte et chacun d'entre eux correspond à une syllabe et couvre une ou plusieurs notes (figure n°3). Par exemple, nous pouvons voir sur la figure n°4 que « torculus »⁶ représente un mouvement mélodique circonflexe bas-haut-médium.

SIGNES NEUMATIQUES à partir du VIIIe	NOTATION CARRÉE à partir du XIIIe	NOTATION MODERNE à partir du XVe
virga		
punctum		
clivis		
podatus		
torculus		
porrectus		

Figure n°4: évolution des symboles utilisés pour la transcription musicale : signes neumatiques à partir du VIIIe, notation carrée à partir du XIIIe et notation moderne à partir du XVe (emprunté à Massin et Massin, 1983 : 129).

⁵ Schuyesmans (2002), adresse de la page : http://www.schuyesmans.be/gregoriaans/FR/FRmu_05.htm

⁶ « Le **Torculus** se compose de trois notes, la médiane étant toujours la plus élevée. Les intervalles entre la première note et la plus haute et entre celle-ci et la suivante peuvent varier » (Schuyesmans, 2002).

Cette spatialisation des hauteurs permet de symboliser la hauteur relative des notes et donc, de transcrire les intervalles. Elle a notamment permis de visualiser ce qui relevait jusqu’alors du « perçu » et donc d’établir une correspondance entre aigu et haut, grave et bas. Mais dans ce genre de notation, les hauteurs des sons ne sont pas encore fixées, en raison de l’absence d’une « hauteur référence ». Ce point de référence apparaît avec la notation diastématique (de « *diastéma* », intervalle), dans laquelle une ligne horizontale indique la hauteur de la note fa. A partir de cette ligne qui est l’ancêtre de notre portée actuelle, il est donc possible de déterminer la hauteur des autres sons (figure n°5). Dans le système de notation mis au point par Guido d’Arezzo (995-1050) en 1025, le nombre de ces lignes « référence » va se multiplier facilitant ainsi la lecture ; ces lignes sont espacées par des tierces et les lignes au-dessus desquelles se trouvent un demi-ton sont colorées (do et fa).

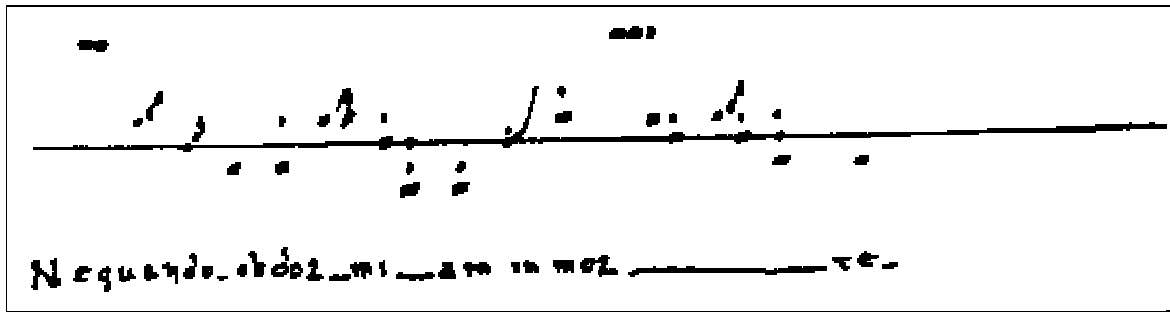


Figure n°5 : notation diastématique, manuscrit Paris B.N. 903 de Saint Yrieix (emprunté à Schuyesmans, 2002)⁵

A la fin du IXe siècle, la notation alphabétique s’associe aux neumes, de manière à fixer les hauteurs (figure n°6) encore plus solidement. La notation neumatique va ensuite peu à peu disparaître pour être remplacée par la notation carrée utilisée par les troubadours à partir du XIIIe siècle (figure n°4). Celle-ci va à son tour céder progressivement la place à la notation mesurée au cours du XIIIe siècle, qui permet de fixer avec une grande précision les rapports de durée entre les notes, précision nécessaire à la coordination des différentes voix dans la polyphonie.

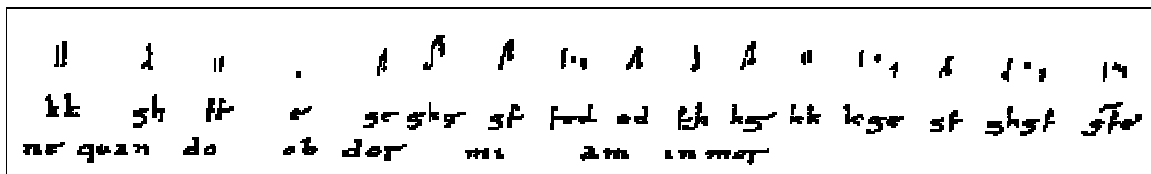


Figure n°6 : double notation musicale, alphabétique et neumatique - Manuscrit de Montpellier H 115 (emprunté à Schuyesmans, 2002)⁶

L'avènement de l'imprimerie au cours du XVe siècle va contribuer à la standardisation de la notation musicale (figure n°4) :

« les procédés de gravure influencent désormais la forme des figures de notes : de carrée ou en losange, elle devient arrondie » (Massin et Massin, 1983 : 133).

Elle se stabilise au cours des XVIe et XVIIe siècle, avec l'apparition de la barre de mesure. Pendant l'époque baroque, la notation garde encore beaucoup de flexibilité du fait de la place centrale accordée à l'interprète :

« jusqu'au XVIIIe siècle, la flexibilité de la notation et, consécutivement, du jeu musical, ne laisse aucun doute ; l'importance des ornements, d'un mode d'écriture comme la basse chiffrée, d'un mode d'expression vocale comme le récitatif nous le confirme incontestablement ; de même, cette surprenante déclaration de Couperin : « Nous écrivons une chose et en jouons une autre » » (Massin et Massin, 1983 : 136).

Cependant, peu à peu, c'est la fidélité au texte qui va s'imposer et la notation se caractérisera par une grande rigidité de la transcription de la hauteur et de la durée. C'est cette transcription musicale « classique » que nous utilisons encore aujourd'hui et qui est enseignée dans les écoles de musique et les conservatoires. Elle n'a plus guère évolué depuis le milieu du XVIIIe siècle. Selon Leipp (1971 : 82) :

« cette notation s'avère extrêmement efficace lorsqu'il s'agit d'instruments traditionnels : il est vraisemblable qu'on ne puisse plus guère l'améliorer. Elle contient la quantité d'information maximale que l'exécutant puisse intégrer pour réaliser en temps réel le programme qui lui est imparté. Elle possède le caractère de ce qui est classique : on ne peut rien en supprimer et l'adjonction d'autres signes la rend confuse, illisible, impraticable ».

La hauteur y est tellement fixée qu'un compositeur comme Bartók a dû introduire un nouveau symbole musical pour représenter les effets de glissement que l'on trouve dans les mélodies populaires (figure n°7) :

« Dans les mélodies populaires [...], il y a beaucoup de sons étrangers, certains glissements de voix, des sons dont la hauteur ne peut être exactement précisée » (cité par Massin et Massin, 1983 : 143).



Figure n°7 : notation du glissando en musique (emprunté à Massin et Massin, 1983 : 143).

1.2.2.3. Des précautions à prendre

Dès lors, nous voyons bien qu'il est difficile de tracer une frontière nette entre la langue et la musique et ce petit détour par l'histoire de la notation musicale montre qu'il faut se munir de beaucoup de précautions quand on essaie de les comparer. Comme le souligne Schaeffer (1966 : 288), il faut se méfier en premier lieu de l'écriture qu'on retrouve dans la note musicale « *en nous la faisant considérer comme un signe préexistant à sa réalisation* ». La transcription « classique » n'est pas la musique, nous l'avons déjà souligné, mais de plus, elle ne permet pas de transcrire toutes les musiques. N'oublions jamais qu'elle se réfère à une **tradition musicale spécifique**, fortement ancrée au niveau historique et géographique. Et il n'est pas envisageable d'aborder sérieusement les similitudes et les différences entre la langue et la musique, sans en tenir compte. Nous voyons peu à peu se dessiner le vide qui existe en la matière et une véritable comparaison entre langue et musique nécessiterait une étude musicologique et ethnomusicologique approfondie, couplée à une étude linguistique et ethnolinguistique comparative. Cette entreprise dépasse largement le cadre de ce travail et nous nous contenterons seulement de préciser les précautions à prendre. La première d'entre elles serait donc de toujours définir la tradition musicale dans laquelle on se place.

1.2.2.4. Tension tonale dans la musique et dans la langue

Selon Menezes (1993 : 84) :

« le fait d'intoner, c'est-à-dire d'être-dans-le-ton se présente, en tant que loi de l'intonation, comme expression de la pensée et de l'émotion aussi bien dans le langage verbal que dans le langage musical. La distinction entre les deux s'avère à ce moment-là impossible à être réalisée face à la « tension tonale » commune à ces deux sphères de la communication humaine ».

Cette tension tonale que nous retrouvons à la fois dans la langue et la musique donne vie à la mélodie. En effet, la mélodie n'est pas une simple succession de sons dans le temps, chacun ayant le même poids que le précédent et que le suivant et la notion de successivité n'implique pas la notion d'égalité : les sons ne sont pas égaux entre eux et certains ont plus d'importance que d'autres. Plus exactement, ils ont plus de poids au niveau perceptif.

Au niveau psychologique, nous sommes sensibles à cette hiérarchie et nous attendons les sons « importants » avec plus d’impatience : cette attente fait naître en nous une sorte de tension qui se résout après que ces sons se soient produits et lorsque l’attente est contrariée, il se crée un effet de surprise. Ces **schémas de tension et de détente** structurent notre perception de la mélodie et sont en grande partie responsables des émotions que nous ressentons à son écoute. Dans la musique tonale, cette hiérarchie porte un nom, la **hiérarchie intra-tonale** :

« à l’intérieur des tonalités, la théorie musicale définit des hiérarchies : la note qui donne son nom à la tonalité (la tonique) est hiérarchiquement la plus importante » (Bigand, 1994 : 254) ;

elle est considérée comme la note la plus stable et les mélodies finissent typiquement sur cette note.

« Puis vient la cinquième note de la gamme, la dominante qui, comme son nom l’indique, domine les autres » (Bigand, 1994 : 254).

De la mélodie émerge l’harmonie puisque un accord de dominante (V^e degré) crée des attentes de résolution sur l’accord de tonique (I^e degré). Du point de vue mélodique, les notes de l’accord de tonique, c’est-à-dire les première, troisième et cinquième notes (voir figure n°8) sont intériorisées mentalement comme les notes les plus stables (Dowling et Harwood, 1986).

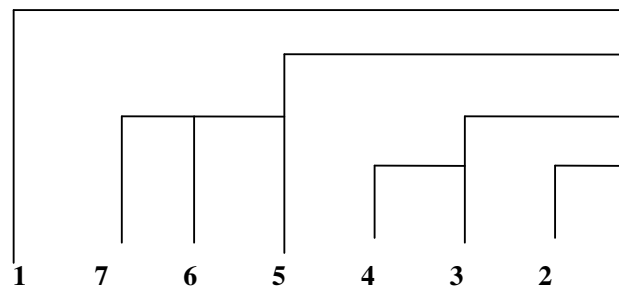


Figure n° 8 : le système de hiérarchie intra-tonale dans la mélodie (d’après Bigand, 1994 : 254).

Par exemple, une mélodie qui aboutit sur la septième note tendra à se résoudre sur une note structurellement plus importante, c’est-à-dire sur la cinquième. Tant que cette résolution ne sera pas accomplie, nous aurons l’impression perceptive de non aboutissement. Arrivé sur la cinquième note, la tension sera toujours présente, mais moins intense. Elle s’apaisera totalement sur la première note, point d’aboutissement de la mélodie.

Les notes peu importantes constituent des points de passage sur lesquels la mélodie ne s'arrête pas. Pour cette raison, on parle de notes de passage. Elles sont notamment utilisées dans les ornements, tels que les appoggiatures, les mordants, les gruppetto et les trilles où la mélodie s'attarde avant de se fixer sur des notes structurellement plus importantes. Elles instaurent de fortes tensions qui seront résolues par l'arrivée de notes plus stables.

Même les adultes qui n'ont pas de formation musicale ont une connaissance tacite des différentes fonctions des notes de la gamme et cette connaissance implique qu'ils soient capables d'attribuer un poids spécifique à chacune des notes au sein d'une mélodie tonale. Ces rapports entre les notes sont intériorisés chez les auditeurs adultes : dans une expérience très connue, Krumhansl et al. (1979) présentent à des sujets musiciens et non musiciens les sept notes de la gamme de Do majeur, puis une des douze notes de la gamme chromatique. La tâche consiste à évaluer sur une échelle de sept points le degré de cohésion d'une note isolée avec la gamme. Pour tous les auditeurs, les notes jugées comme s'accordant le moins bien avec la séquence sont les notes extérieures à la tonalité de Do majeur. Par contre, la tonique do est jugée comme la note la plus intégrée à la séquence. Chez les sujets les plus musiciens, la dominante, puis la tierce sont les notes jugées les mieux intégrées après la tonique. Ces rapports de hiérarchie semblent intériorisés de manière très précoce : en effet, dès 4 mois et demi, les bébés préfèrent des extraits tirés des menuets de Mozart segmentés correctement (Jusczyk, 1989), car ils écoutent plus longtemps les passages interrompus par des pauses situées aux limites des phrases plutôt qu'à l'intérieur des phrases. Une phrase non résolue du point de vue tonal ne peut pas être considérée comme achevée. Bien sûr, d'autres paramètres peuvent expliquer cette préférence pour les « bonnes limites », tels que la cohérence temporelle et rythmique de la phrase, mais du point de vue perceptif, la hiérarchie tonale semble également jouer un grand rôle. Outre le fait que cette hiérarchie entre les notes permette à l'auditeur de structurer sa perception de la mélodie et par conséquent, de l'harmonie, elle lui donne la faculté de prédire les notes qui vont se produire. Grâce à l'intériorisation de ces schémas, nous sommes en effet capables de compléter une ligne mélodique s'il manque une note. Nous sommes également capables de prédire la fin d'une phrase musicale et ce, même pour un air que l'on n'a jamais entendu auparavant. Selon Jackendoff (1991), l'auditeur anticipe ce qui est à venir (« *prospective hearing* ») en fonction de ce qu'il a déjà perçu (« *retrospective hearing* »).

La musique dodécaphonique a remis en cause ce type de hiérarchie tonale en proposant des gammes où chaque demi-ton possède un poids équivalent à celui des autres demi-tons. Mais on peut se demander si au niveau perceptif, l'auditeur de ce type de musique ne continue pas à analyser les sons de manière hiérarchique, en raison de son acculturation au système tonal et ce, depuis l'enfance (voir chapitre II, 3.1.3. et chapitre III, 2). Pour Lerdahl (1985), l'absence de hiérarchie dans les musiques contemporaines non tonales rend plus difficile la structuration perceptive et la représentation intériorisée des œuvres. Selon lui,

« les systèmes compositionnels actuels tendent à être inaccessibles, même à des auditeurs expérimentés » (1985 : 110).

Mais, cette inaccessibilité ne proviendrait-elle pas tout simplement du manque d'« acculturation » aux musiques non tonales ? Après tout, la musique tonale résulte d'une tradition musicale spécifique, la musique occidentale classique (et notamment, la musique tonale où la mélodie est soumise à l'harmonie, contrairement à la musique modale grégorienne). Cette tradition s'est construite au cours des siècles, jusqu'à atteindre son apogée pendant la première moitié du XIXe siècle. Elle a ensuite évolué naturellement vers l'atonalité, à cause de la saturation du système tonal (voir 2.2.2.). Il est vrai qu'aujourd'hui, la musique diffusée est majoritairement tonale, notamment dans les musiques de variété et qu'il est donc impossible de se soustraire à ses influences. Elle est en quelque sorte notre « *langue maternelle musicale* » (Viret, 1987 : 7). Mais, cela ne veut pas pour autant dire que le système tonal soit plus naturel que les autres systèmes, comme certains auteurs l'écrivent. Nous en prendrons pour preuve l'acculturation avec des systèmes musicaux qui nous paraissent beaucoup plus complexes, tels que les gamelans balinaïses (voir chapitre II, 3.1.3.3.).

Existe-t-il un système de hiérarchie similaire dans la langue ? Il existe dans la parole des points « cibles » hauts et des points cibles bas, étrangement fixes chez un même locuteur. Si le locuteur s'interrompt en plein milieu d'une phrase, la courbe mélodique de l'énoncé se trouve comme suspendue et attend la résolution, que ce soit sur un point cible haut comme dans la question ou sur un point cible bas, comme dans l'affirmation. Cette attente crée un fort sentiment de tension chez l'auditeur. Par exemple, « Je vais au marché ↑ » [...] « demain ↓ ».

Cet effet de tension peut être utilisé en rhétorique, en particulier en insérant des pauses là où on ne les attend pas (cf. les discours du général De Gaulle) : le locuteur réussit à « suspendre » l'auditoire à ses lèvres et il est fort probable que la mélodie soit déterminante dans ce phénomène.

Ainsi, les facultés de prédiction qui se manifestent pour la musique se retrouvent également dans la langue : nous sommes tous capables de compléter des phrases qui ne sont pas terminées (nous finissons la phrase de notre interlocuteur ou nous parlons en même temps). Dans cette prédiction, nous nous basons sur des indices syntaxiques et sémantiques, mais avant tout mélodiques, comme le prouvent les travaux de Grosjean à ce sujet. Selon Grosjean (1983), la mélodie permet en effet de faire ressortir un certain point de l'énoncé appelé « dernier mot potentiel » (DMP), à partir duquel l'auditeur pourrait prédire la fin de la phrase. Les paramètres utilisés dans cette mise en relief, tels que la rupture de hauteur pour la Fo et l'allongement de la durée du DMP, joueraient le rôle de clés prosodiques. Pour savoir si l'information sur la longueur de la phrase en anglais est disponible uniquement sur le DMP, Grosjean et Hirt (1996) ont mené une série de quatre expérimentations. Bien que la syntaxe et la sémantique soient utilisées dans la première partie de la phrase, il arrive un moment où elles n'apportent plus d'information sur la continuation potentielle de la phrase. Cette information, seule la prosodie est en mesure de la donner. C'est ainsi que les sujets réussissent à prédire trois types de phrase lors de la seule écoute du DMP : ils peuvent affirmer que la phrase est terminée, qu'il reste encore trois mots ou six mots.

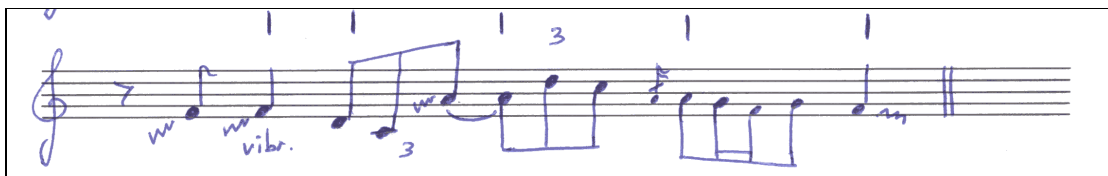
Que se passe-t-il avec le français, une langue dont l'organisation prosodique est très différente de celle de l'anglais ? Là où il existe trois choix possibles pour l'anglais (énoncé terminé ou trois mots ou six mots et plus), il n'en reste plus que deux pour le français (énoncé terminé ou énoncé qui se continue). Les anglais et les français réagissent de manière différente pendant les tâches de prédiction car l'information donnée dans leur langue respective est différente, les clés prosodiques ne les informant pas de la même manière (cf. chapitre V, partie 2.). Des sujets manifestant des lésions de l'hémisphère droit rencontrent de graves difficultés dans l'acte de prédiction, surestimant les phrases se terminant sur le DMP et sous-estimant celles qui ne se terminent pas sur le DMP (cf. chapitre VII, 3, discussion sur l'hémisphère droit, l'« hémisphère musical »).

La mélodie semble donc jouer un rôle déterminant dans le traitement des énoncés, puisqu'elle fournit des informations sur leur gabarit temporel. Il nous est arrivé, en travaillant sur un corpus de bébés de 9 à 48 mois (Dodane et al., 1999), de rencontrer des difficultés à délimiter les différents énoncés. Selon les critères adoptés dans cette étude, nous posons une limite lorsque la pause dépassait une certaine durée (400 ms). Mais, il arrivait de temps à autre que, malgré une pause de grande longueur, nous ayons l'« intuition » que l'énoncé n'était pas terminé. Et seule la mélodie pouvait nous fournir cette information, en assurant la cohésion d'un énoncé dont les parties étaient éloignées dans le temps. En outre, elle contient des indices sur les frontières linguistiques importantes, qui permettent de segmenter les énoncés. Ainsi, de manière frappante, les bébés font preuve des mêmes capacités de segmentation que pour la musique. Dans une étude que nous avons déjà citée, Jusczyk (1989) a montré que des bébés anglophones de 4 mois ½ préféraient des passages interrompus aux limites des phrases plutôt qu'à l'intérieur. Ces résultats ont été confirmés chez des bébés de 6 mois par Hirsh-Pasek et al. (1987) et plus récemment par Nazzi et al. (2000).

1.2.2.5. Différenciation mélodie langue / musique

Comment mettre à jour les caractéristiques spécifiques de la mélodie de la langue par rapport à celle de la musique ? Il paraît difficile de le faire sans effectuer une comparaison acoustique et auditive des deux types de mélodie. Or, il existe un cas de figure qui pourrait nous faciliter la tâche : certains musiciens ont essayé d'imiter l'intonation des langues naturelles avec leur instrument (voir 2.2.3.4.). Jimi Hendrix est peut être le plus célèbre d'entre eux : dans « *Still Raining, Still Dreaming* » (1968), il débute son solo de guitare par une imitation de l'intonation humaine qui dure quelques secondes (2021 ms), puis à la phrase suivante, la mélodie « redevient » musique (**extrait musical n°1**). Puisqu'une guitare est éloignée de la voix humaine, le musicien est obligé de « forcer » le trait, tout comme le ferait un imitateur lorsqu'il imite la voix d'un homme politique ou un dessinateur lorsqu'il réalise une caricature. Pour que son imitation soit réussie, il lui faut exagérer les traits caractéristiques du modèle. C'est à cette condition que l'auditeur pourra reconnaître l'extrait en tant qu'« imitation » de l'intonation. Un tel extrait se montre donc précieux pour nous aider à révéler les différences entre l'intonation et la mélodie, car ces différences y seront forcément exagérées.

Les deux phrases ont d'abord été transcrites sur portée musicale, de manière à visualiser leur rythme aussi bien que leurs notes (figures n°9 et n°11). Leur ligne mélodique a été resynthétisée grâce au logiciel Praat afin d'isoler la guitare des instruments qui l'accompagnent (basse et batterie), ce qui pourrait perturber la détection de la Fo. Une première analyse de leur courbe mélodique nous révèle que la phrase « musicale » (figure n°10 ; **extrait musical n°2**) contient plus de glissements que la phrase « parlée » (figure n°12 ; **extrait musical n°3**) et que les hauteurs semblent moins fixées. Les instruments à corde et notamment la guitare électrique (effets obtenus notamment avec la pédale wah wah dans cet extrait) permettent en effet des effets de glissando, qui sont beaucoup plus difficiles à obtenir avec des instruments « à clés » dont les hauteurs sont fixées (flûtes traversières, saxophones, clarinettes...). Hendrix joue de ces effets de manière beaucoup plus marquée pour la musique que pour l'intonation : il commence sa phrase musicale par un glissando montant de 6 demi-tons (216-336 Hz) qui aboutit au fa³, suivi d'un vibrato centré autour du la³, très visible sur la représentation mélodique (indiqué par une flèche sur la figure n°10). On observe également des glissements pour la phrase parlée (voir figure n°12), notamment une chute vertigineuse de 641 Hz à 97 Hz (soit 33 demi-tons !) au début. Mais, les hauteurs suivantes sont plus fixes, surtout à la fin de la phrase. On voit donc bien que ces glissements ne sont pas l'apanage de la parole et qu'ils sont aussi difficiles à transcrire pour la musique que pour la langue, comme nous l'avons d'ailleurs expérimenté ici (figure n°9 et n°11). Passy reprochait à la notation musicale de se montrer inapte à transcrire la mélodie de la parole, mais il aurait dû préciser qu'il parlait de la notation musicale classique. Et c'est justement le filtre de cette notation qui lui fit supposer que la caractéristique essentielle de la langue par rapport à la musique était de glisser d'une hauteur à une autre. Cependant, si la musique contient plus de glissements dans cet extrait, d'où nous vient alors ce sentiment de « fixité » des hauteurs à son écoute ?



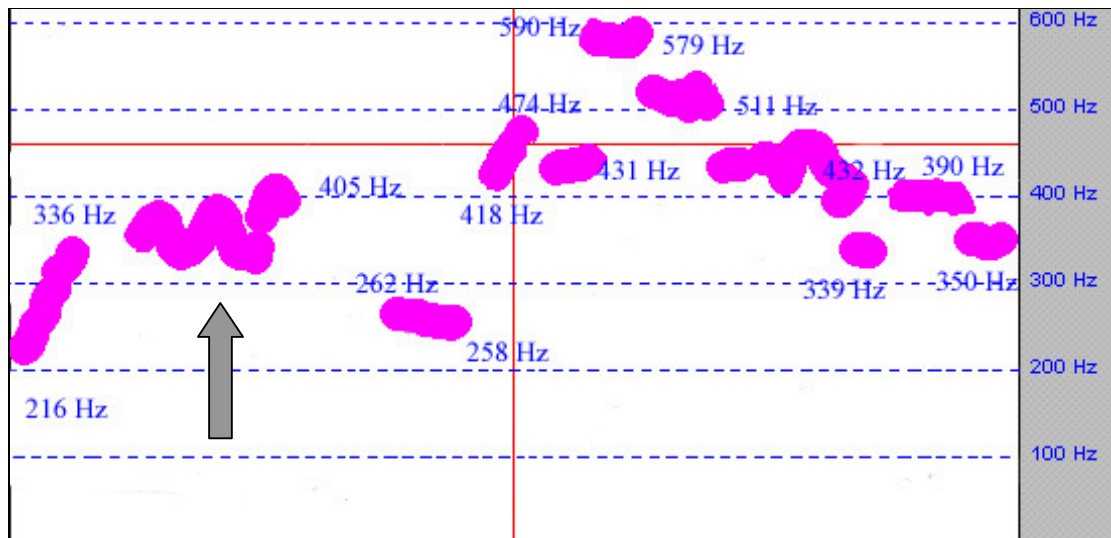


Figure n°10 : représentation de l'évolution de la Fo obtenue après resynthèse avec le logiciel Praat.

La mélodie de la phrase musicale est structurée autour de l'accord parfait Fa M, composé des trois notes structurellement les plus importantes de cet accord : fa (Tonique, I), do (Dominante, V) et la (Médiane, III). Après un glissando qui aboutit sur le fa³, la mélodie reste fixée sur cette note, qui possède le poids le plus important dans la tonalité (tonique). Cette note est allongée (croche liée à une noire, soit 1 temps 1/2) et placée sur un temps fort, qui renforce son poids. Elle est suivie d'un triolet dont la dernière note est le la³, c'est-à-dire la médiane. Grâce à la liaison rythmique, cette note est allongée et positionnée sur la partie forte du temps suivant. La mélodie passe ensuite par le ré (VII), puis par do (V), revient sur le la (III), notes d'ornement, puis fa (I). Elle pourrait s'arrêter sur cette note, mais retarde la résolution en incluant une nouvelle note de passage. La mélodie aboutit sur le fa³, placé sur un temps fort et allongé (noire). Dans cet extrait, les notes tonales importantes occupent donc une position forte (partie forte du temps ou temps fort dans la mesure). Cet « ancrage tonal » est renforcé par un allongement important. Ainsi, la tonique fa (I) et la médiane la (III) sont au minimum deux fois plus longues que les autres notes (voir tableau n°2). Les plus longues valeurs correspondent donc aux « bonnes » notes, c'est-à-dire aux notes tonales importantes. **La mélodie est animée, pour ne pas dire « rythmée » par l'attente et la venue des trois personnages principaux que sont dans un ordre hiérarchiquement décroissant la tonique, la dominante et la médiane.** La structuration rythmique, en réservant aux notes tonales les positions fortes, assure une forte cohésion avec la mélodie. Les « bons intervalles » sont ainsi fixés aussi bien au niveau tonal que rythmique. Par ailleurs, on observe une continuité de la mélodie dans le sens où le plus grand intervalle entre deux notes consécutives est une 6te M (do³-la³).

Note	Fa	Fa	Ré	Do	La	Ré	Do	La	La	Sol	Fa	Sol	Fa
Durée en ms	330	252	172	199	335	160	172	67	301	58	80	243	437

Tableau n°2 : durée des différentes notes composant la mélodie de la phrase "musicale" (en ms).

La mélodie de la phrase « parlée » commence par un mi⁴ pour chuter très rapidement (au cours d'un triolet) jusqu'à un autre mi, le mi², deux octaves plus bas. Cette dernière note est allongée (noire) et suivie d'une pause (demi-soupir), qui délimite une première unité. La mélodie, qui reprend au si³, effectue ensuite une série de sauts d'octave très rapides (au cours d'un triolet) : elle chute d'une octave (pour les correspondances des intervalles musicaux en demi-tons, voir p : X) sur un autre si (si²), puis remonte d'une octave pour atteindre un do (do⁴), puis un do# (do#⁴) et chute à nouveau de plus d'une octave jusqu'au sol (sol²), qui est allongé et suivi d'une pause. Cette dernière note marque la fin d'une seconde unité et la fin de la phrase « parlée ». Les deux pauses servent donc à délimiter deux groupes au sein de la phrase.

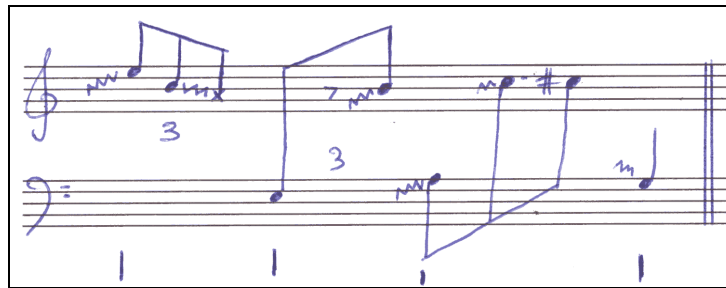


Figure n°11 : transcription musicale de l'extrait de « Still Raining, Still Dreaming » où Jimi Hendrix fait « parler » sa guitare.

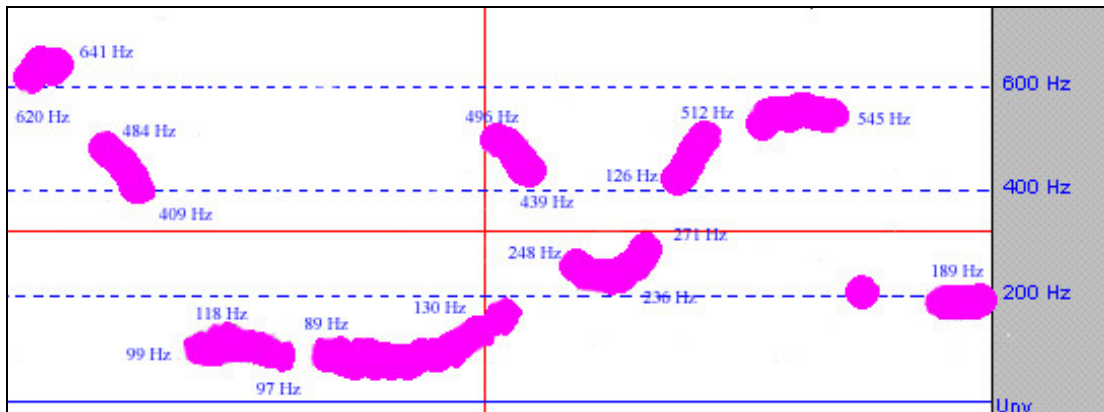


Figure n°12 : représentation de l'évolution de la Fo obtenue après resynthèse avec le logiciel Praat.

A l'intérieur de chaque groupe, la durée des différentes notes est sensiblement égale, ce qui donne à l'auditeur l'impression qu'elles s'écoulent régulièrement (tableau n°3). Il n'y a que la dernière note du deuxième groupe qui soit plus allongée (277 ms), mais le rapport avec les autres notes n'atteint jamais le double.

Note	Mi	Si	La	Mi	/	Si	Si	Do	Do#	Mi
Durée en ms	213	162	175	172	255	172	232	191	178	277

Tableau n°3 : durée des différentes notes composant la mélodie de la phrase « parlée » (en ms).

Aucune note à l'intérieur des groupes n'est mise en relief par un allongement de la durée, comme les notes tonales importantes pour la musique. La mélodie du premier groupe commence et se termine par un mi ; celle du second groupe commence par un si, qui est répété et aboutit sur un sol. Autrement dit, les notes importantes sont mi-sol-si, soit un accord parfait mineur. Nous n'irons pas plus loin, mais ces notes ne sont pas mises en relief aussi fortement que ne l'étaient les notes de l'accord parfait pour la musique.

Par ailleurs, on peut observer une grande discontinuité dans la ligne mélodique ; la mélodie franchit souvent l'intervalle d'octave entre notes successives (plus grand intervalle : 8ve + 5te A, soit sol2-do#4), alors que pour la phrase musicale les mouvements étaient conjoints. Selon Vos et Troost (1989), la plupart des types de mélodie utilisent typiquement des changements de hauteur de l'ordre de 1/12 à 1/6 d'octave et la subdivision de l'octave à travers les cultures tend à se placer dans cette étendue. Dans la parole, les variations de hauteur tendent à dépasser la moitié de l'octave et les relations entre les intervalles ne sont pas critiques (Patel et al., 1998). Outre leur grande étendue, ces changements de hauteur se montrent continus à l'intérieur des groupes : il n'y a jamais d'arrêt sur une note particulière qui serait allongée, et de plus, on observe une succession rapide de montées et de descentes.

Ces caractéristiques semblent être les plus importantes pour différencier la mélodie de la langue et celle de la musique dans cet extrait : les écarts entre les notes sont trop grands et les notes importantes du point de vue tonal ne sont pas marquées au niveau du rythme. Plus la mélodie est rangée et plus les « bonnes » notes sont mises en relief, plus la mélodie est perçue comme musicale, plus les intervalles sont fixés (car on laisse à l'oreille le temps de les fixer ; avec la parole, la succession de notes est trop rapide). La cohésion mélodique et rythmique de la phrase musicale assure la « fixation » des intervalles et donc de la tonalité. C'est donc cette propriété qui différencie l'intonation de la musique, et non pas les glissements d'une note à l'autre.

1.2.3. A mi chemin entre musique et langue : les clichés mélodiques

Les clichés mélodiques (on parle également d'« intonations clichés » ou, en anglais, de « stylised tunes ») sont constitués par une séquence de mots dont la courbe mélodique est fortement contrainte, voire figée.

Ces clichés sont communs aux membres d'une même communauté et utilisés dans des situations bien spécifiques (Di Cristo, 1998 : 215), telles que les salutations par exemple (« *Coucou* » et « *Bonjour* » ; « *hello* » ou « *bye bye* ») ou le fait de répondre au téléphone (« *allô* » ; « *moshi moshi* » en japonais). On les retrouve en abondance dans le discours des commissaires-priseurs, des vendeurs et notamment sur les marchés, ... et ils sont très courants dans la parole de l'enfant, surtout dans la raillerie (« *Na-na-na-na-nère* »). Ils trahissent la langue maternelle des locuteurs qui s'expriment dans une langue étrangère et contribuent ainsi fortement à l'accent d'intonation (Fonagy et al., 1976).

Les clichés mélodiques sont très intéressants car ils révèlent la proximité de la langue et de la musique : ils se situent dans une position intermédiaire entre la parole et le chant (Hirst, 1998 : 71). Il est donc particulièrement pertinent de mettre à jour ce qui contribue à faire d'un cliché un motif musical, afin de pouvoir commencer à esquisser une frontière entre la langue et la musique.

Les clichés mélodiques se caractérisent par une stabilité de hauteur sur chaque syllabe (régularité micro-prosodique) :

« instead of consisting of a continuous sequence of movements from one target point to the next as in normal speech, the contour is produced as a sequence of static level tones » (Hirst, 1998 : 71).

Cette régularité contribue à donner au cliché son aspect chantonné (Fonagy et al., 1983 : 153). Elle permet de fixer les intervalles intersyllabiques et assure ainsi une régularité au niveau macro-prosodique. La transcription musicale est particulièrement apte à schématiser les clichés mélodiques (Léon, 1971 : 124), car elle fait apparaître cette régularité et cette stabilité au niveau de la distribution des hauteurs. Selon Fonagy :

« la mélodicité des énonciations constitue, à côté de la hauteur et du temps, une troisième dimension de l'espace imaginaire où évolue la mélodie » (Fonagy, 1983 : 155).

Dans l'exemple du cliché « *na-na-na-na-nère* » (voir transcription musicale, figure n°13), la ligne mélodique est tellement figée qu'elle peut être représentée par une série de notes musicales : sol-sol-mi-la-sol-mi, qui correspond à la succession d'intervalles : tierce mineure (sol-mi), quarte juste (mi-la), seconde majeure (la-sol) et enfin tierce mineure (sol-mi).

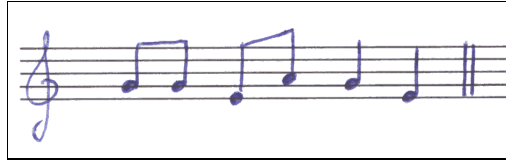


Figure n°13 : transcription musicale du cliché mélodique « na-na-na-na-nère »

C'est cette succession d'intervalles qui permet à l'auditeur de mémoriser et de reconnaître l'énoncé « na-na-na-na-nè-re », plus que le contenu segmental qui n'a d'ailleurs aucun sens. En anglais, on retrouve le même cliché avec le même contour mélodique, mais des syllabes légèrement différentes, « Na-na-na-na-na-euh ». Le pouvoir évocateur des clichés est souvent utilisé par les musiciens, notamment dans les musiques improvisées, car il suffit de reproduire le contour du cliché pour que les auditeurs le reconnaissent immédiatement. Dans « *The Jump* », le chanteur Bobby McFerrin (1984) incorpore le cliché musical « Na-na-na-na-na-euh » à une phrase musicale (**extrait musical n°4**). Cette citation est très intéressante, car bien que le texte soit différent (il utilise les syllabes « le-le-le-le-le-le »), nous identifions tout-de-suite le cliché mélodique correspondant. En revanche, dans la deuxième phrase musicale (voir transcription musicale, figure n°14), il modifie le contour en produisant un fa bécarré au lieu d'un fa dièse (soit un demi-ton plus bas), c'est-à-dire un intervalle de tierce majeure au lieu d'une tierce mineure, puis une quarte augmentée au lieu d'une quarte juste (pour les non musiciens, se reporter au tableau des correspondances p : 7). Bien qu'il contribue à l'expressivité de la phrase musicale, ce « petit » changement suffit à briser le cliché. Les intervalles ne souffrent donc aucun changement et la forme est complètement figée. Par ailleurs, la rythmique est régulière et facilement restituable grâce à la notation musicale :



Figure n°14 : extrait de « *The Jump* » de Bobby McFerrin (1984).

1.2.4. Degré de mélodicité et de rythmicité, la clé de la musicalité ?

D'après Fonagy (1983 : 155) :

« c'est le degré de mélodicité plus ou moins élevé qui distingue les différents genres du discours : chant, récitatif, sermon, déclamation, discours politique, exposé, conversation ».

La mélodicité semble donc bien être un critère essentiel pour caractériser une parole musicale et donc pour la distinguer d'un discours « non musical ». D'après nos observations, le degré de mélodicité pourrait se définir par le degré de fixation des notes tonales importantes, par l'intermédiaire de leur allongement et/ou de leur emplacement au sein de la cellule rythmique. Le degré de rythmicité participe également à la musicalité d'un énoncé. Nous le définirons comme le degré de stabilité de la rythmique. Si ce degré est fort, la pulsation sera facilement reproductible par l'auditeur. Nous pourrions formaliser ces relations de la manière suivante : un axe vertical borné en haut par l'« intonativité » et en bas par la « musicalité ». Plus le degré de mélodicité et plus le degré de rythmicité seront importants, plus l'énoncé considéré sera musical. On peut donc classer les différents types de discours langagiers et musicaux sur ce même axe. C'est ce que fait implicitement Van Waesberghe (1957 : 374, cité par Léon et Martin, 1969 : 66-67) lorsqu'il classe les différents types de communication, de la parole normale à la musique. Sur la figure n°15, nous avons fait correspondre l'axe de musicalité à cette typologie : ainsi, la parole normale possède la plus forte « intonativité » alors que le chant avec rythme imposé possède la plus grande « musicalité ».

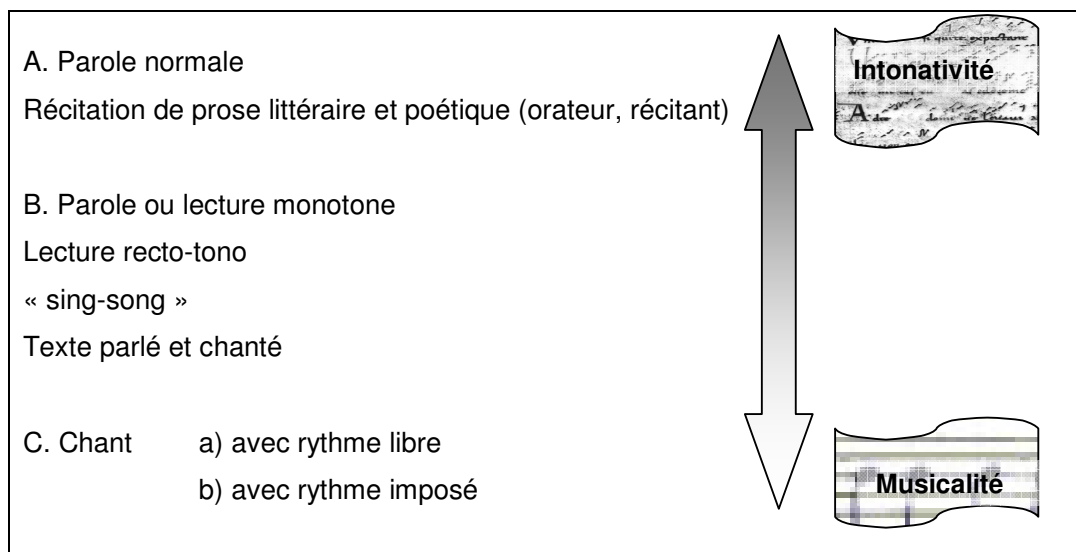


Figure n°15 : typologie des différents types de communication par Van Waesberghe (1957 : 374, citée par Léon et Martin, 1969 : 66-67).

Il nous semble donc beaucoup plus pertinent de classer les différents types d'énoncés en fonction de leur musicalité, c'est-à-dire de leur mélodicité et de leur rythmicité, plutôt qu'en fonction de leur appartenance à la musique ou à la langue.

1.3. Cas de fusion avec la musique : les langues « musicales »

Dans certaines langues, le message ne passe plus que par les éléments musicaux, comme c'est le cas dans les langues sifflées et les langues tambourinées. Ces langues « musicales » imitent le rythme, l'intonation et les tons de la langue parlée correspondante, par l'intermédiaire du sifflement ou par l'intermédiaire d'un instrument de musique. Ce procédé d'imitation est très semblable à l'effet de « guitare parlante » utilisé par Hendrix (cf. 1.2.2.5. et **extrait musical n°3**). C'est la musique de ces langues, c'est-à-dire sa prosodie qui est utilisée pour faire passer des messages normalement véhiculés par le contenu segmental. En l'absence de mots, l'ambiguïté est forcément plus grande, mais reste toutefois limitée en raison de la restriction des situations de communication pendant lesquelles les messages sont échangés. Dans ces langues, on assiste à une véritable fusion de la prosodie et de la musique.

1.3.1. Langues sifflées

Les langues sifflées sont utilisées pour communiquer sur de grandes distances. Grâce à une bande de fréquence étroite (2 KHz), la portée du signal permet en effet de couvrir aisément plusieurs kilomètres (jusqu'à 12 à 14 km pour le Silbo). On a recensé jusqu'à présent quatre langues sifflées. Elles sont la plupart du temps utilisées par des montagnards, et spécialement les bergers qui peuvent échanger des messages d'une vallée à une autre, en dépit du relief accidenté. Elles sont également utilisées dans la conversation courante. La radiocinématographie montre que, pendant la production d'un énoncé sifflé, les organes articulatoires bougent de la même façon que pour la phrase parlée équivalente. Selon les langues, le sifflement peut suivre le pattern du deuxième formant du langage ordinaire du locuteur (Shadle, 1997 : 53). Celui-ci est quasiment réduit à une fréquence pure qui imite les inflexions de l'intonation et les caractéristiques de la langue parlée correspondante. La langue sifflée des Indiens d'Oaxaca au Mexique est fondée sur le mazateco, qui est une langue tonale (Cowan, 1948, cité par Crystal, 1987 : 400). Chaque syllabe correspond à une syllabe de la parole et les 4 tons de la langue parlée sont utilisés. Ces tons sont classés de 1 (haut) à 4 (bas). Cowan transcrit la phrase équivalente en français à « Qu'est-ce que tu m'apportes là ? » de la façon suivante : 1,1,3,3,2,4 hme¹ č²a¹ š³ ki³ -č²ai²-ve⁴.

Dans les îles Canaries, on trouve une autre langue sifflée, le Silbo de la Gomera qui est fondée sur un dialecte de l'espagnol (Busnel et Classe, 1976). Sur les bords de la Mer Noire, une autre langue sifflée est pratiquée dans le village de Kusköy en Turquie. Cette langue s'appuie sur le turc. Enfin, la dernière langue sifflée a été recensée en France, dans le village d'Aze situé dans les Pyrénées. Elle est fondée sur un sous-dialecte du béarnais, dont la matrice est espagnole.

1.3.2. Langues tambourinées

Les phénomènes rythmiques peuvent également porter le message linguistique : c'est le cas des fameux « tambours parlants » utilisés dans certains pays d'Afrique. Tout comme le sifflement, l'utilisation des percussions permet de délivrer des messages relativement complexes et sur de longues distances. Les langues tambourinées s'appuient sur l'imitation du rythme de la langue parlée correspondante, de manière plus ou moins codifiée. Mais elles peuvent également imiter des phénomènes mélodiques, c'est-à-dire l'inflexion de la voix et les tons. Par exemple, les membres de la tribu Jabo, à l'Est du Libéria utilisent des troncs de 2 mètres de long, dont la fente longitudinale permet de produire différents tons (Herzog, 1945). Ces instruments permettent non seulement de reproduire la rythmique du Jabo, mais également d'en reproduire les 4 tons, et ce qui lie ces tons entre eux, les glissandos. Sur la transcription réalisée par Herzog (figure n°16), la rythmique est indiquée par la valeur des notes (croches, double-croches) et des silences (demi-soupir), c'est-à-dire leur durée. Les tons sont indiqués par les chiffres affectant les différentes syllabes (de 1 à 4), comme c'était le cas pour le mazateco (cf. 1.3.1.).

The image shows three staves of musical notation for the Jabo language. Each staff consists of a musical staff with notes and rests, followed by syllables with tone markers (superscripts) and a French translation. The first staff has two measures: the first measure contains the syllables 'nā⁴' and 'wi¹ε¹o²' with the translation 'Greetings!'; the second measure contains 'ba²' and 'di²²le¹' with the translation 'Come ye quick! Put ye your effort there!'. The second staff contains the syllables 'cla²wle¹', 'Gb²nā⁴', 'ba²', 'te²', 'Zle²le²', 'ba²', 'b²', 'd³', and 'd³' with the translation 'Soldiers all! Stop ye the noise. Speak ye one by one!' (Played in the men's assembly when the discussion threatens to get out of hand.). The third staff contains the syllables 'Du²i²', 'bl²ε³', 'ka²nō²⁻¹', 'Gwe²nē¹', 'a³', 'mī²', 'Du²i²', 'bl²ε³', and '(c¹)' with the translation 'To collect fines – hunger is raging – we are going to collect fines (indeed)' (Played before the assembly sends out a group to collect fines imposed by the court.).

Figure n°16 : transcription du Jabo (emprunté à Crystal, 1987 : 400).

Dans sa thèse, Essien (2000) décrit le fonctionnement des tambours parlants yoruba utilisés au Nigéria. Des tendeurs relient les deux membranes situées à chaque extrémité de l'instrument. Il est possible de faire varier la tension de ces tendeurs en exerçant une pression variable à l'aide du bras. Cette variation continue de la tension sur la membrane va engendrer une variation continue de la hauteur, lorsqu'on martèle rapidement la peau avec une baguette. Grâce à ce système, le tambour est donc « capable » de reproduire les glissements de l'intonation de la langue naturelle.

Essien explique que :

« des personnes dans l'audience, qui connaissent la mélodie et la chanson, peuvent substituer les sons de la parole aux hauteurs et entendre la « parole » dans la mélodie jouée » (Essien, 2000 : 63).

Ils ont dû au préalable passer par une phase d'apprentissage qui leur a permis d'intérioriser des phrases de la langue naturelle et leurs correspondances tambourinées. La langue yoruba se prête bien à cette « musicalisation », car selon Beier (1976, cité par Essien, 2000 : 63), les tons (trois tons, bas, moyen et haut) sont plus significatifs que les voyelles et les consonnes. Ainsi, beaucoup de mots se distinguent uniquement par leurs tons et non par leur contenu segmental. Pour cet auteur, la capacité du tambour à produire ces tons, équivaut donc à la capacité à parler. Et cette capacité a toujours fasciné les voyageurs, comme le souligne Essien. Il cite un extrait d'*Awake* (1997) qui raconte la transmission d'un message :

« In May 1976, the remarkable ability of drummers to communicate using drums was demonstrated by court musicians of a Yoruba chief. Volunteers from the audience whispered a series of instructions to the master drummer who, in turn, drummed the instruction to another musician located far from the courtyard. Responding to the drum instructions, the musician moved from one location to another and carried out whatever action he was requested to perform » (cité par Essien, 2000 : 62).

1.3.3. Une langue artificielle musicale

S'il existe des langues qui utilisent uniquement les éléments musicaux comme vecteurs linguistiques, en se fondant principalement sur les patterns rythmiques et intonatifs de la langue correspondante, alors, pourquoi ne pas imaginer une langue artificielle qui serait fondée uniquement sur le langage musical ?

Une telle langue pourrait avoir des prétentions universelles, puisque le langage musical « classique » et la notation qui lui est associée sont connus par un grand nombre de personnes à travers le monde. C'est ainsi qu'au début du XIXe siècle, le compositeur français Jean-François Sudre (1787-1862) créa une « langue musicale universelle », une sorte d'espéranto musical baptisé « solrésol ». Ce projet fut soutenu entre autres par Victor Hugo, Alexandre Humboldt et Alphonse de Lamartine et fut très populaire au XIXe siècle, notamment à la cour de Napoléon III. Il constitue l'une des plus vieilles tentatives connues de création d'un langage artificiel. En 1902, Borelas Gajewski en rédigea la grammaire. Le solrésol est fondé sur les 7 notes de la gamme diatonique : **do, ré, mi, fa, sol, la** et **si** (transcrites **d, r, m, f, so, l, s**). Ces notes sont utilisées comme des syllabes et elles peuvent se combiner entre elles pour créer des mots. Produites de manière isolée, la note **si** signifie « oui » et la note **do**, « non ». La combinaison de deux notes engendre des mots grammaticaux, tels que « je » (**dr**) et « tu » (**dm**) ou d'autres mots tels que « bonsoir ; bonne nuit » (**mso**). La combinaison de trois notes est utilisée pour les mots courants, tels que « boire » (**rsol**), « jour » (**drm**), « du pain » (**dsom**), « mai » (**dll**), « soleil » (**soff**) et « siècle » (**drs**). Les mots de quatre notes sont rangés par classes, identifiées par la note initiale. Ainsi,

« les mots commençant par SOL ont rapport au théâtre, à la littérature, aux beaux-arts et aux sciences » (Gajewski, 1902),

somsor pour « musique », par exemple. Il y a en outre plus de 9000 combinaisons à 5 notes pour les noms d'animaux, de végétaux et de minéraux. Le solrésol permet d'obtenir des oppositions sémantiques simplement en inversant les syllabes : par exemple **sofs** pour « rire » et **fso** pour « pleurer ». On peut indiquer le genre (indiqué par une barre horizontale : « femme », **dmfr** ⁻) et le nombre (indiqué par un accent aigu : « mesdames », **ss'o** ^ˆ). Les verbes sont au participe présent, le temps verbal étant indiqué par une lettre (**D** pour l'imparfait de l'indicatif par exemple : « je commençais », **dr D sdf**). Chaque mot est suivi d'un bref silence ce qui permet à l'auditeur de bien les délimiter.

Le solrésol a été conçu pour pouvoir couvrir un éventail très large de situations de communication. Ainsi, il peut se transcrire avec plusieurs systèmes différents : en toutes lettres, avec des notes sur une portée musicale (la portée utilisée comporte seulement trois lignes, car la gamme utilisée s'arrête au **do**⁴), avec les 7 premiers chiffres ou des couleurs spécifiques ou des frappes rythmiques associés aux 7 notes. Il peut même être sténographié. Par ailleurs, il peut être indifféremment sifflé, chanté, parlé ou joué à l'aide d'un instrument.

Comme pour les langues sifflées et les langues frappées, il peut donc se transmettre sur de longues distances, de jour comme de nuit. En dépit de ses qualités, le solrésil n'a pourtant jamais atteint l'universalité à laquelle le prédestinait son concepteur. Aujourd'hui, très peu de personnes communiquent avec ce langage. Il nécessite en effet un long apprentissage, car les combinaisons sont difficiles à mémoriser (en raison du nombre limité d'unités). Dans sa version, chantée, sifflée ou jouée, il nécessite de la part du locuteur et de l'auditeur une très bonne oreille musicale, voire l'oreille absolue pour repérer la note initiale des mots de 4 notes. Par ailleurs, si l'auditeur est habitué à attribuer un sens bien précis à des combinaisons de notes en solrésil, il ne pourra plus s'abstenir de le faire lorsqu'il écoutera de la musique. Il est probable qu'il perde le plaisir de la musique, la musique étant pansémique. En plus d'une bonne oreille, il faut chanter juste en solrésil pour ne pas brouter le message. En réalité, le solrésil n'a de musical que le nom, car il utilise des principes très différents de la musique. Alors que :

« d'une manière tout à fait générale, on pourrait dire que, parmi toutes les manières possibles de structurer, à des fins quelconques, le continuum sonore, celles qu'utilisent la musique et le langage se situent à des niveaux différents et peuvent être réalisées simultanément sans qu'il se crée d'interférences » (Ruwet, 1972 : 50).

Pourvu que la fonction distinctive soit assurée dans le langage, il existe une grande liberté dans l'utilisation matériel sonore à des fins musicales. Mais, le problème du solrésil, c'est que le niveau du langage est justement attribué à la musique : les combinaisons de notes jouent le rôle des mots dans les langues naturelles, exactement comme si on utilisait la prosodie à la place des mots. La fonction du contour mélodique n'est pas de porter un sens aussi précis que ne le font les mots. Par ailleurs, dans sa version écrite, le solrésil n'est ni plus, ni moins qu'un langage de signaux semblable au morse. Ce n'est donc pas les similitudes entre langue et musique qui sont exploitées ici.

1.4. L'origine de la langue dans la musique ?

Les similitudes très fortes entre langue et musique et notamment entre la langue parlée et la langue chantée ont poussé certains linguistes, compositeurs et philosophes à poser l'hypothèse d'une origine commune du langage et de la musique. Cette idée se retrouve très fréquemment au cours du XVIIIe siècle. Ainsi, selon Rameau :

« on peut dire que la musique, simplement considérée dans les différentes inflexions de la voix, laissant le geste à part, a dû être notre premier langage » (cité par Didier, 1982 : 400).

Mais selon Kintzler (citée par Didier, 1982 : 400), chez Rameau :

« langue et musique n'ont pas la même identité, elles sont simplement en conformité acoustique, en tant qu'émettrices de vibrations sonores ».

Alors qu'il fonde cette origine uniquement sur les similitudes acoustiques entre la langue et la musique, la question des origines se place sur un terrain essentiellement philosophique pour Condillac et Rousseau. Tandis que Condillac, dans son *Essai sur l'Origine des Connaissances Humaines* (1746) évoque une époque où langage, musique et geste n'auraient fait qu'un, Rousseau, dans son *Essai sur l'Origine des Langues, où il est Parlé de la Mélodie et de l'Imitation Musicale* (1763) se plaît à imaginer la nature des premiers discours des êtres humains, où musique et langue n'étaient pas différenciées selon lui :

« Autour des fontaines dont j'ai parlé les premiers discours furent les premières chansons : les retours périodiques et mesurés du rythme, les inflexions mélodieuses des accens firent naître la poésie et la musique avec la langue, ou plutôt tout cela n'étoit que la langue [...] Les premières histoires, les premières harangues, les premières loix furent en vers ; la poésie fut trouvée avant la prose ; cela devoit être, puisque les passions parlèrent avant la raison. Il en fut de même de la musique ; il n'y eut point d'abord d'autre musique que la mélodie, ni d'autres mélodie que le son varié de la parole, les accens formoient le chant, les quantités formoient la mesure, et l'on parloit autant par les sons et par le rythme que par les articulations de la voix. Dire et chanter étoit autrefois la même chose dit Strabon » (1990 : 114).

Pour Condillac, la musique et la langue se sont ensuite progressivement séparées, selon un processus d'une extrême lenteur :

« la prosodie s'éloigna [...] du chant peu à peu, et à mesure que les raisons qui l'en avoient fait approcher davantage cessèrent d'avoir lieu. Mais elle fut longtemps avant de devenir aussi simple qu'elle l'est aujourd'hui » (1746 : 119, chapitre II, première section, seconde partie « De la Prosodie des Premières Langues »).

Avant cette séparation, la parole donnait sens à la musique, mais la variation musicale pouvait également changer le sens d'un mot. On relève d'ailleurs ce phénomène dans les langues tonales où le changement de ton est distinctif. Ainsi, Condillac apporte pour preuve de l'origine commune de la langue et de la musique l'existence d'une langue comme le chinois, car dans cette langue, la variable qui fait sens peut être d'ordre musical (le ton) ou d'ordre linguistique (le phonème) :

« dans l'origine des langues, la manière de prononcer admettoit donc des inflexions de voix si distinctes qu'un musicien eût pu la noter, en ne faisant que de légers changemens ; ainsi je dirai qu'elle participoit du chant. Cette prosodie a été si naturelle aux premiers hommes qu'il y en a eu à qui il a paru plus facile d'exprimer différentes idées avec le même mot, prononcé sur différents tons, que de multiplier le nombre des mots à proportion de celui des idées. Ce langage se conserve encore chez les Chinois. Ils n'ont que 328 monosyllabes qu'ils varient sur cinq tons, ce qui équivaut à 1640 signes » (Condillac, 1746 : 119, seconde partie, section première, chapitre II « De la Prosodie des Premières Langues »).

Pour Rousseau et Condillac, c'est au moment de la naissance de l'écriture, que la séparation entre la langue et la musique s'est amorcée. Près de deux siècles plus tard, le musicologue soviétique Assafjew (1884-1949) explique ce processus de différenciation de la manière suivante (cité par Menezes, 1993 : 87-88) : certains patterns mélodiques se sont peu à peu fixés et leur reconnaissance grâce à la mémoire auditive a permis la constitution d'une conscience musicale collective. C'est alors que la scission avec le langage s'est réalisée. La musique s'est notamment spécialisée en développant une véritable science des intervalles :

« La précision de plus en plus incisive de l'intervalle en tant qu'indicateur de la tension propre à l'intonation et l'évolution assez frappante des structures musicales dans l'histoire envisageant surtout la composition des intervalles constituent sinon la raison d'une séparation absolue, au moins une distinction assez nette entre musique et langage verbal (ou poésie). [...] L'harmonie, science des proportions d'intervalles, émerge, nous le constatons, en tant qu'élément le plus éminent du discours musical – l'intonation mélodique en constituant sa manifestation la plus élémentaire –, et l'intervalle sera vu comme une forme primordiale de la musique en tant qu'expression autonome, bien qu'il ait irréfutablement, et justement pour cette raison, son origine au sein même du Verbe » (Menezes, 1993 : 87).

Selon Menezes (1993 : 113), cette spécialisation de la langue et de la musique se serait manifestée par une différenciation des écritures verbales et musicales. L'écriture alphabétique a omis, avec l'établissement de l'alphabet, la transcription de l'intonation et des relations de durée des phonèmes dans leurs réalisations, en excluant la prosodie de son contexte. L'avènement plus tardif de l'écriture musicale est dû aux relations d'intervalles et de durées, développées de façon toujours plus systématique et exhaustive que dans le langage parlé. L'écriture musicale est donc née en raison de la spécialisation du code musical vis-à-vis des propriétés prosodiques du langage et a contribué à la démarcation historique des limites entre codes musical et linguistique.

Enfin, parmi d'autres théories, Jespersen (1922 : 401-402) cite celle de Noiré qui situe l'origine du langage dans le rythme : lors de tâches effectuées en commun, les hommes produisaient des grognements rythmés pour accompagner l'effort physique qu'ils fournissaient. Ces sons finissaient par s'associer à l'acte lui-même et servaient à le désigner :

« en conséquence, les premiers mots signifieraient quelque chose comme « hisser » ou « haler » » (Jespersen, 1922 : 402).

Cette théorie est surnommée la « yo-he-ho » theory ou si nous la traduisons en français, la théorie « eh-oh-eh-oh ». Jespersen refuse de spéculer sur l'origine du langage et utilise un raisonnement inductif fondé sur l'observation du langage enfantin, du langage des peuplades primitives et de l'histoire du langage. Comme Rousseau et Condillac, il en tire la conclusion que le premier langage se rapprochait plus du chant, que de la parole telle que nous la connaissons aujourd'hui :

« et il nous faut donc conclure que le langage des hommes non civilisés et primitifs contenait bien plus d'agitation passionnée que le nôtre et ressemblait plus à un chant, une musique » (Jespersen, 1922 : 406).

Selon Jespersen, l'origine du langage doit se rechercher en effet dans les aspects poétiques et romantiques de l'existence et notamment, dans l'expression des émotions :

« Le langage naquit à l'époque des amours de l'humanité ; et je veux croire que les premiers énoncés produits tenaient à la fois de la plainte nocturne d'une chatte amoureuse sur un toit et des mélodieux chants d'amour du rossignol » (Jespersen, 1922 : 419-420).

Dans ce langage « primitif » fondé exclusivement sur l'expression des émotions, la langue et la musique ne se différenciaient pas, elle formaient un tout. Ainsi, pour Jespersen, *« il a existé une époque où tout langage était chant, ou plutôt durant laquelle ces deux choses n'étaient pas encore distinctes »*. Il ajoute :

« L'homme a chanté ses sentiments bien avant d'être capable d'exposer ses pensées. Mais il ne faut pas, bien entendu, nous imaginer que « chanter » a dans ce contexte la même signification que dans une salle de concert moderne. Lorsque nous disons que la parole est née du chant, ce que nous voulons dire, c'est essentiellement que notre langage relativement monotone et que notre musique vocale très élaborée ne sont que les résultats de la différenciation d'énoncés primitifs, qui ressemblaient moins au premier qu'à la seconde » (1922 : 422).

2. Exploitation de ces similitudes de forme

Tout au long de ces lignes, nous avons vu que les rapports étroits entre la langue et la musique puisaient leurs origines dans les similitudes très fortes qui existaient entre la musique et la prosodie. Mais, d'après Menezes (1993 : 113), il existe un véritable paradoxe dans le fait que le lien entre la musique et la langue ne puisse se faire que par l'intermédiaire de la parole. En effet, depuis l'avènement de l'alphabet, la prosodie a été complètement exclue de la transcription orthographique ; celle-ci ne permet absolument pas de rendre compte de l'intonation et des rapports de durée entre les phonèmes. La prosodie appartient donc uniquement à la sphère de l'oral, c'est-à-dire à la parole. C'est pour cette raison que les linguistes et en particulier les phonéticiens ont dû créer des systèmes de notation de la prosodie pour combler le vide laissé par l'orthographe (cf. partie 2.1.1.2. et 2.1.2.). Au contraire, l'écriture musicale s'est spécialisée sur la transcription de la prosodie, notamment sur les relations entre les hauteurs et entre les durées des sons. Elle n'a pas cessé de s'affiner, jusqu'à atteindre une très grande précision dans la notation des intervalles, dans la fixation des hauteurs et des durées (voir historique de la notation musicale, partie 1.2.2.2. et 2.1.1.1.). C'est grâce à son écriture que la musique a pu différencier les niveaux acoustique, instrumental et stylistique, et par conséquent, qu'elle a pu exercer un contrôle absolu de chacun de ces niveaux. Les propriétés acoustiques ont ainsi pu se libérer des conventions auxquelles elles étaient soumises. L'écriture musicale a permis également d'approfondir les rapports entre les propriétés acoustiques et leurs propriétés auditives. Elle a donc permis l'évolution des méthodes de composition et l'évolution du style musical. C'est ce dynamisme qui distingue la notation musicale de la notation orthographique, cette dernière possédant un caractère éminemment immuable. Pour Menezes (1993 : 119) :

« l'écriture musicale est née en raison de la spécialisation du code musical vis-à-vis des propriétés prosodiques du langage et a corroboré de façon décisive la démarcation historique des limites entre les codes musical et linguistique ».

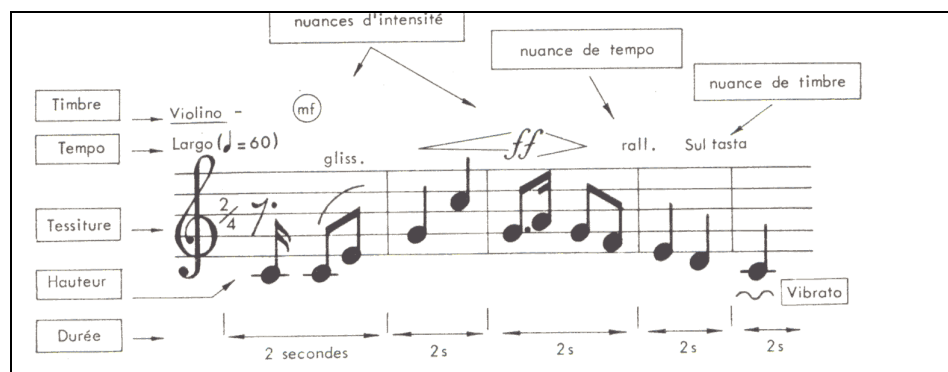
2.1. Lorsque la musique devient langue

2.1.1. Transcription musicale de la prosodie

2.1.1.1. Potentialités de la transcription musicale

Nous avons vu dans le paragraphe 1.2.2. que la mélodie de la musique et celle de la parole ont beaucoup de similitudes, notamment dans le fait qu'elles se caractérisent par une succession de sons au cours du temps (dimension temporelle) et que ces sons entretiennent entre eux des rapports de hauteur (intervalles), des rapports de durée, mais aussi des rapports d'intensité (ils ne sont pas tous prononcés avec la même force articulatoire). On relève également des similitudes entre le rythme de la musique et celui de la langue, dans le fait qu'ils se définissent tous deux par une succession de constituants forts et de constituants faibles. Ces constituants forts sont mis en relief par une combinaison particulière des trois paramètres acoustiques que sont la durée, la Fo et l'amplitude. Cette mise en relief correspond au phénomène d'accentuation, présent à la fois dans la langue et dans la musique. A la question de savoir comment représenter la mélodie et le rythme d'un énoncé, les phonéticiens se sont donc tout naturellement tournés du côté de la transcription musicale classique, qui permet justement de représenter la durée, la hauteur et l'intensité (voir figure n°17). Selon Leipp (1971 : 81),

« cette « partition » si simple et, point important, lisible de façon quasi instantanée, comporte, en fait, un nombre extraordinaire de renseignements sur les variables physiques et psychophysiques en présence ».



A la lumière du parallèle langue / musique, les indications qui s'avèrent particulièrement intéressantes pour la transcription la prosodie portent sur⁷ :

- **les indications de durée** : l'**axe horizontal** de la portée qui représente la dimension temporelle de la langue et de la musique ; les **valeurs des notes** (la noire dure un temps par exemple) **et des pauses** (la demi-pause dure deux temps par exemple) pour la durée des syllabes et des pauses dans la parole ; l'indication de vitesse d'exécution, c'est-à-dire le **tempo**, pour le débit (noire=60, 60 noires par minute ; on trouve d'autres indications de tempo qui permettent d'accélérer ou de ralentir au cours de l'exécution, comme « *accelerando* » ou « *rallentando* ») ; les indications qui concernent la mesure (le chiffrage 2/4 indique qu'il y a deux noires par mesure) et les barres de mesure (barres qui permettent de délimiter les mesures) semblent inappropriées, car elles impliquent une trop grande métricité, absente dans la langue. Il paraît donc difficile de faire tenir le rythme de la langue entre deux barres de mesure.
- **les indications de hauteur** : l'**axe vertical** de la portée qui représente l'étagement des hauteurs du plus grave au plus aiguë (de bas en haut), pour indiquer les rapports de hauteur entre les sons (notes et syllabes) ; la **clef** (ici la clef de sol & permet de fixer une référence en associant une hauteur absolue, la note sol³, 392 Hz, à la deuxième ligne de la portée) et les **cinq lignes de la portée** permettent de fixer la **hauteur** (une fois la référence fixée grâce à la clé, les lignes de la portée permettent de représenter les hauteurs en fonction de ce point) ;
- **les indications d'intensité** : symboles ajoutés en dessous de la portée (***pp*** pour « *pianissimo* », très doux ou ***ff***, pour « *fortissimo* », très fort par exemple) ; certains d'entre eux indiquent une augmentation (« *crescendo* ») ou une diminution progressive de l'intensité (« *decrescendo* »), voir figure n°18.

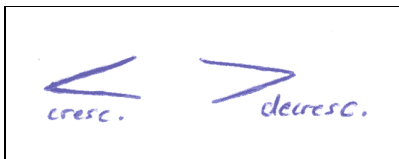


Figure n°18 : symboles utilisés dans la notation musicale pour indiquer une augmentation progressive de l'intensité (*crescendo*) ou une diminution progressive de l'intensité (*decrescendo*).

⁷ Pour plus de détails sur la notation musicale elle-même, se reporter à Leipp (1971 : 81-82).

2.1.1.2. Historique des applications de la transcription musicale en prosodie

Les potentialités de l'écriture musicale ont séduit nombre de linguistes et de phonéticiens et ce, depuis de nombreux siècles. Il y a eu beaucoup de tentatives d'adaptation de cette notation aux particularités de la prosodie des langues naturelles et il est possible d'en retracer rapidement l'historique en s'arrêtant aux noms les plus connus. Ces tentatives se heurtent aux obstacles des différences entre langue et musique ou réussissent au contraire à les dépasser. Elles permettent de compléter les différences et les similitudes mises à jour dans la partie 1.2.

Dès le XVI^e siècle, le grammairien Louis Meigret, cité par Konopczynski (1978 : 82) tente de représenter l'évolution de l'intonation et de l'accentuation sur une portée musicale dans son *Treuvé de la Gramèze françoèze* (1550). Son système n'avait recours qu'à une seule valeur de note, la blanche et se limitait à deux hauteurs différentes. Une transcription aussi primitive n'exploitait que très partiellement les possibilités de la notation musicale et ne pouvait bien sûr pas rendre compte de la richesse musicale d'une langue naturelle, tant du point de vue de ses variations de hauteur, que de ses durées. Près de deux siècles après Meigret, Joshua Steele fit une exploitation beaucoup plus approfondie de la notation musicale dans une étude portant sur l'intonation anglaise, « *Prosodia Rationalis or An Essay Towards Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols* » (1775)⁸. A l'opposé de celle de Meigret, la transcription de Steele est extrêmement précise dans la notation des durées et de la hauteur. Faure reproche d'ailleurs à Steele ce degré de précision :

« l'auteur s'efforce d'autre part, ce qui est fort aléatoire, [...] de noter des différences de hauteur de l'ordre du quart de ton, du demi-ton et de trois quarts de ton entre les syllabes successives d'un énoncé et des repos allant jusqu'au huitième de pause » (1962 : 118).

En réalité, cette précision s'explique par le procédé adopté par Steele pour transcrire la mélodie (Crystal, 1987 : 170) : il a indiqué sur le manche d'une viole de gambe (basse de viole, même tessiture que le violoncelle) les hauteurs correspondant aux **frets** (à droite de la figure n°19). A la manière d'Hendrix, il a ensuite reproduit les variations de hauteur de la voix sur le manche de la viole et les a transcrites sur une portée musicale (à gauche de la figure n°19). La transcription présentée est celle d'un extrait de « *Hamlet* » de Shakespeare, le célèbre « *To be or not to be...* ». La durée de chacun des glissement est symbolisée par des signes verticaux, indiquant une durée relative (du plus court au plus long).

⁸ Cité par Faure (1962 : 18), Crystal (1969 : 23 ; 1987 : 170) et Konopczynski (1978 : 82).

Le degré d'emphase figure en dessous de la portée ; les symboles sont également relatifs. Les pauses sont signalées par des symboles empruntés à la notation musicale : les silences.

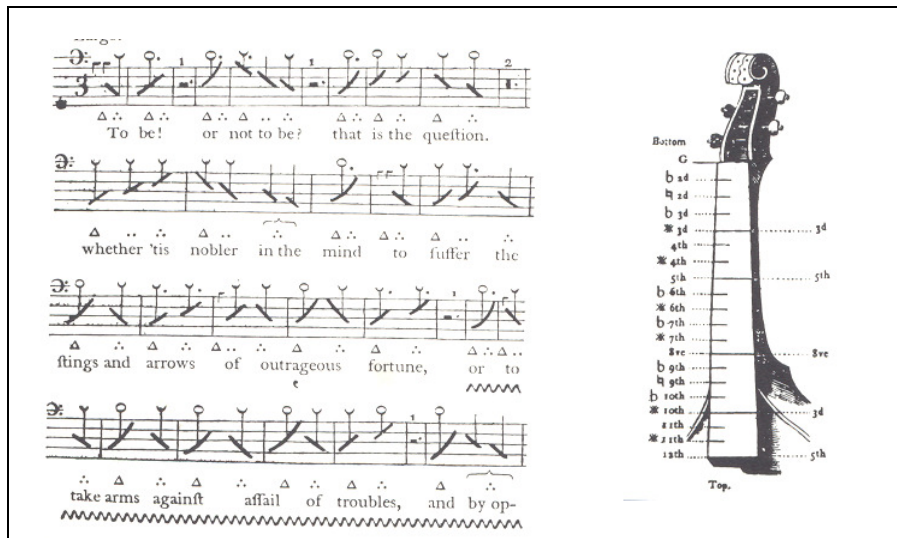


Figure n°19 : à gauche, premières lignes de la transcription utilisée par Steele ; à droite, manche de la viole de gambe tel qu'il a été utilisé par Steele (emprunté à Crystal, 1987 : 170).

Steele se rendit compte que les points de départ et d'aboutissement des glissandi se trouvaient très souvent entre deux frets et ne correspondaient donc pas aux notes de la gamme diatonique. Pour cette raison, ces points sont la plupart du temps localisés entre deux lignes de portées sur sa transcription. Un siècle plus tard, le français Léonce Roudet (1899) préconisa une adaptation de la notation musicale, car il avait remarqué lui aussi que les hauteurs ne correspondaient pas aux hauteurs des notes de la gamme. Pour cela, il utilisa une échelle logarithmique en Hz, sur laquelle étaient indiquées les correspondances avec les notes de la gamme. Ces notes étaient prolongées par des lignes horizontales qui remplaçaient la portée traditionnelle (10 lignes, étagées de mi^2 -165 Hz à sol^3 -396 Hz, soit une tessiture d'une dixième mineure). En outre, pour Roudet, la notation musicale s'avérait inapte à représenter les variations continues de hauteur. Pour palier ce problème, il les représenta donc par une ligne continue évoluant sur sa portée à 10 lignes. Les pauses étaient indiquées par une rupture de cette ligne mélodique (figure n°20).

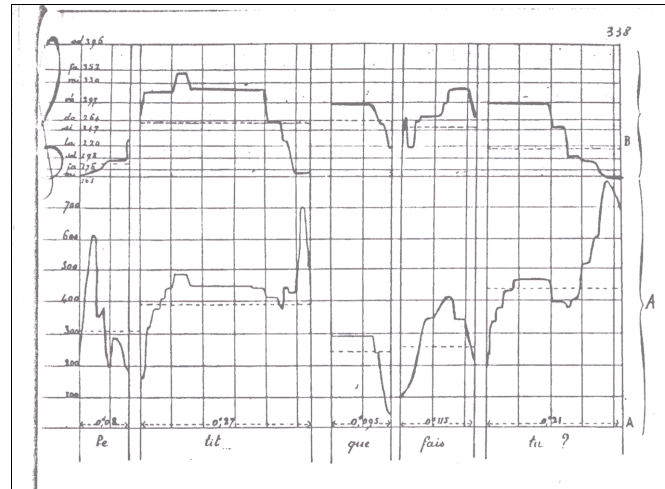


Figure n°20 : transcription utilisée par Roudet (1899),
emprunté à Konopczynski, 1978 : 93

L'axe horizontal représentait l'axe temporel comme dans la notation musicale. Mais la durée ne pouvait plus être indiquée par les valeurs des notes, ni par les barres de mesures, car elles avaient été supprimées par Roudet. L'axe temporel était donc gradué selon l'échelle suivante : 1 cs = 1,6 mm (Konopczynski, 1978 : 84). Quant à l'intensité, elle était représentée par une autre courbe, placée juste en dessous de la courbe mélodique. Le système de notation de Roudet ressemble beaucoup aux représentations des courbes mélodiques et des courbes d'amplitude que l'on obtient avec les logiciels d'analyse phonétique modernes. L'abbé Rousselot (1901-1908) utilisera, quelques années plus tard, le même type de transcription à quelques différences près. L'échelle utilisée pour les hauteurs n'est plus logarithmique, mais musicale et graduée par demi-tons. Une courbe d'amplitude accompagne la courbe mélodique, mais elle figure au-dessus d'elle (figure n°21). Quant à la durée, l'unité de mesure n'est pas précisée car Rousselot reproduit les enregistrements.

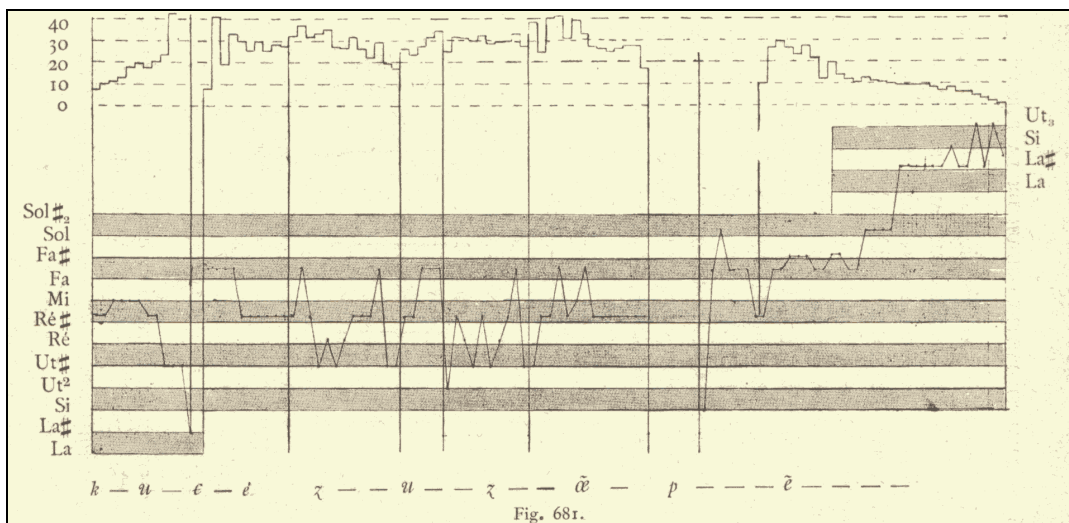


Figure n°21 : transcription utilisée par Rousselot (1901-1908 : 1005).

Ce système de transcription est donc nettement moins précis que ne l'était celui de Roudet. En 1909, paraissent les célèbres « *Intonation Curves* » de Daniel Jones. Dans cet ouvrage, Jones utilise le système de notation musicale pour transcrire la prosodie de textes dramatiques dans deux langues différentes : l'anglais et le français. Contrairement à Roudet et Rousselot, il ne modifie pas la portée musicale, mais comme eux, il représente l'évolution de la hauteur dans la parole par une courbe continue (continuité de la ligne mélodique anglaise). Cette courbe est fondée sur des bases auditives. Il est à souligner que le contour de la mélodie est interrompu à l'emplacement des sourdes (sons non voisés, pas de hauteur). La transcription phonétique (sons et accents) du texte figure en dessous de la portée et les différentes syllabes sont délimitées par des barres de mesures (voir figure n°22, empruntée à Konopczynski, 1978 : 95). La durée de chaque syllabe est représentée par la longueur relative de la mesure. En dehors de ces barres de mesure, il n'y a pas d'indications de durée. On ne trouve pas non plus d'indications d'intensité. Cette transcription est donc beaucoup plus simple que les trois précédentes, mais surtout beaucoup plus lisible. C'était le but recherché par Jones, dont l'approche était avant tout celle d'un didacticien. Les phonéticiens Roudet et Rousselot visaient au contraire à transcrire avec le plus de détails possibles l'évolution des paramètres acoustiques et ce, dans un but d'analyse. Et il est normal d'adapter le système de transcription à ses objectifs. De manière générale, tous ces auteurs ont essayé d'adapter le système de notation musicale, afin qu'il reflète les spécificités de la musique des langues naturelles.

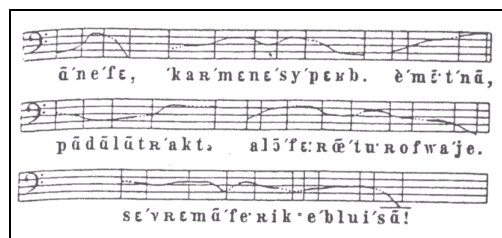


Figure n°22 : transcription utilisée par Jones dans « *Intonation Curves* » (1909)
(emprunté à Konopczynski, 1978 : 95).

Le système de transcription adopté par le psychologue allemand Wilhelm Wundt est un exemple intéressant, car dans un premier temps, cet auteur a transposé la notation musicale dans le domaine de la langue sans la modifier. Selon Konopczynski (1978 : 88),

« Wundt est si influencé par la musique qu'il en arrive à assimiler la phrase à un air dont il commente les caractéristiques ».

Il utilise même une armure, ce qui implique une tonalité. Steele et Roudet avaient justement abandonné les notes musicales, car elles symbolisaient des hauteurs fixes, correspondant à la gamme diatonique. Avec ce système le plus petit intervalle possible entre deux notes est le demi-ton (gamme tempérée) et les intervalles sont toujours décomposables en un nombre de demi-tons entier (multiples). On rencontre ce type d'intervalles dans la langue (voir tableau d'intervalles), mais également des intervalles avec des rapports plus « complexes ». Un des principaux dangers avec la notation musicale, c'est d'interpréter ces intervalles complexes en fonction de la structure des « bons » intervalles du système tonal, que nous avons intériorisés depuis l'enfance. Cela nous rappelle étrangement le phénomène de « crible » phonologique, qui nous pousse à interpréter les phonèmes d'une langue étrangère en fonction de la structure des phonèmes de notre langue maternelle (cf. chapitre V, partie 1.2). Un musicien, lorsqu'il transcrit la mélodie de la langue, aura tendance à « normaliser » les intervalles et sa transcription sera faussée. On pourrait parler de crible « tonal » et il est fort probable qu'il s'applique à la perception des musiques non tonales (cf. chapitre II, 3.1.3.3.). Wundt s'est vite rendu compte qu'il fallait abandonner la rigidité des notes musicales en raison de cette

« différence essentielle entre parole et musique, cette dernière contenant toujours des intervalles harmoniques ; d'où la difficulté d'appréciation des intervalles de hauteur, généralement rapprochés des intervalles musicaux connus, et donc minimisés » (Konopczynski, 1978 : 89)

et que pour palier ce phénomène de crible, il était nécessaire de coupler une analyse instrumentale à l'analyse auditive.

En 1900, il adopte un nouveau système de transcription fondé sur une analyse instrumentale et dans laquelle les variations mélodiques sont représentées par une ligne continue, qui évolue sur une portée musicale à 5 lignes (figure n°23). Des notes sont ensuite affectées aux points d'inflexion de la mélodie. Les accents sont indiqués sur la transcription orthographique qui figure au-dessus de la portée (accent aigu et accent grave). Quant à la durée, elle est représentée par l'espacement relatif entre les différentes notes. Signalons la grande ressemblance avec le système de Jones.



Figure n°23 : transcription utilisée par Wundt (1900)
emprunté à Konopczynski, 1978 : 97.

Mis à part Rousselot et Roudet, tous les auteurs évoqués jusqu'à présent avaient conservé la portée musicale à cinq lignes. Pike (1945) sera le premier à bouleverser ce système en introduisant la notion de « niveaux » intonatifs (on parle également de registre). La tessiture qui correspond à une portée musicale (de re^3 à sol^4 , soit un intervalle de 17 demi-tons, 293,67 Hz-783,99 Hz) ne correspond pas à la tessiture couverte par la voix dans la parole. Pour complexifier encore les choses, cette tessiture n'est pas la même selon les langues. Elle est par exemple beaucoup plus grande en anglais (plus de deux octaves, voir chapitre V, partie 1.1) qu'en français. Il faut donc adapter la notation musicale à la tessiture de la langue étudiée. Pour répondre à ce problème de tessiture, il est possible de diviser l'espace fréquentiel couvert par la voix en un certain nombre de zones, appelées registres ou niveaux, relatifs bien sûr au registre de l'individu. Pike (1945) fut le premier à employer les niveaux intonatifs, dans la description de la prosodie de l'anglais américain. De 5 lignes, la portée passe à 4 lignes, chacune représentant un niveau distinct ; ces niveaux sont espacés par un même intervalle de hauteur. La portée est divisée horizontalement en intervalles de temps égaux ; pauses indiquées par la rupture de la ligne mélodique. Pour l'intensité, on trouve l'indication de la place et de la force de l'accent. Ce système de niveaux a été complété par Faure (1962) pour la description de l'anglais britannique. Cette langue se caractérise en effet par de fortes variations mélodiques qui lui donnent des propriétés très musicales (voir chapitre V). Faure est passé à un système à 5 niveaux, ce qui redonne à la portée le même « visage » que la notation musicale.

Ce système à 5 niveaux a été repris par Zwanenburg (1965) et Delattre (1969) pour le français (voir figure n°24). La notation de Delattre est essentiellement mélodique, elle ne comporte aucune indication de durée et d'intensité.

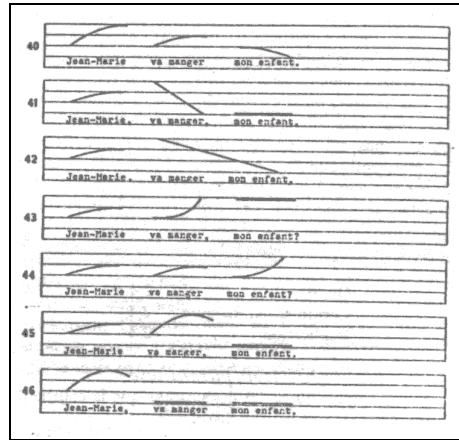


Figure n°24 : transcription de la phrase « Jean-Marie va manger mon enfant » par Delattre (1969 : 12).

Léon et Léon (1964) ont également utilisé une représentation par niveaux, mais au lieu de représenter la courbe mélodique, ils ont inséré une transcription orthographique de l'énoncé dont les différentes syllabes se répartissent sur quatre niveaux différents. Ce procédé est similaire à celui de Bolinger (voir plus loin, figure n°31), qui fait « chanter » les lettres en leur faisant suivre l'évolution de la mélodie, à la différence près que celui-ci ne se réfère pas à la notion de niveau et qu'il n'a pas recours à une « portée ».

Nous l'avons vu, toute transcription répond à des objectifs bien précis. Et lorsque le but de la recherche est justement de comparer, voire de rapprocher la langue de la musique, une transcription musicale authentique paraît la mieux appropriée. Fonagy et Magdics (1963) n'ont pas eu recours aux niveaux, car leur objet d'étude porte sur la comparaison des émotions dans la parole et dans la musique vocale européenne. Bien au contraire, l'utilisation de la notation musicale leur permet de comparer la ligne mélodique de la langue à celle de la musique. Ils sont donc restés très fidèles à la transcription musicale, comme on peut le voir sur la figure n°25.



Figure n°25 : extrait de la transcription de Fonagy et Magdics (1963).

permettre de visualiser facilement l'intonation en même temps que le rythme, un système de transcription a été mis au point (figure n°27) : l'énoncé, transcrit orthographiquement figure dans une large flèche dont les courbes suivent l'évolution de la ligne mélodique. Cette flèche vient se superposer à une portée musicale. Les syllabes dont les caractères sont indiqués en gras doivent être accentués.



Figure n°27 : transcription musicale utilisés dans la méthode « L'Anglais par le Rap » (emprunté à Ribière-Raverlat, 1997 : 185).

Depuis la blanche de Meigret jusqu'aux niveaux de Delattre ou Faure, il a fallu toute une évolution historique pour que la notation musicale s'adapte à la « musique » de la langue. Toutes ces tentatives ont permis d'approfondir progressivement les connaissances sur les différences et les similitudes entre la langue et la musique. Par ailleurs, l'apparition des analyseurs de mélodie a permis de résoudre les problèmes liés à l'analyse auditive pure, qui concernaient principalement les erreurs d'appréciation des intervalles par l'oreille à cause du crible tonal. Mais à l'inverse, l'analyse instrumentale ne doit pas faire perdre l'usage de la notation musicale, car celle-ci obéit justement à une analyse auditive qui manque évidemment à l'analyse instrumentale. Le danger est donc de perdre l'ancrage auditif de la mélodie et donc son ancrage musical. Le couplage des deux types d'analyses paraît donc être un très bon compromis, comme Wundt l'avait déjà compris il y a plus d'un siècle. Les nouveaux systèmes de notation de la prosodie (cf. partie 2.1.2.) ne doivent pas sonner le glas de la transcription musicale. Chaque système répond en effet aux objectifs d'analyse bien spécifiques de son utilisateur. Des chercheurs comme Fonagy et Magdics, Papousek et Papousek ont utilisé la notation musicale car celle-ci constituait le système le mieux adapté à leurs objectifs de recherche. Massin et Massin (1983 : 124) faisaient remarquer à propos des différents types de transcription utilisés pour la musique que

« chercher un système de notation généralisable à toutes les musiques correspondrait à la volonté d'imposer un unique système de pensée et d'analyse ».

Cette remarque vaut tout autant pour les systèmes de transcription de la prosodie et elle a d'autant plus de résonance que la tendance actuelle en matière de transcription va justement vers une uniformisation des pratiques.

2.1.1.3. Transfert de la terminologie musicale vers la prosodie

Reflétant l’analogie entre langue et musique, de nombreux termes musicaux ont été transférés du domaine de l’analyse musicale à celui de la description prosodique. Il n’est qu’à se référer aux glossaires de deux livres parus récemment et concernant la prosodie du français : « *La Prosodie du Français* » de Lacheret-Dujour et Beaugendre (1999 : 271-288) et « *L’Intonation, le Système du Français : Description et Modélisation* » de Rossi (1999 : 203-212). Dans le premier ouvrage, on trouve 16 mots provenant du domaine musical : accent, ambitus, anacrouse, attaque, cadence, coda, gamme tonale, glissando, intervalle mélodique, octave, pauses, registre, tempo, tessiture et unité tonale. Dans le second ouvrage, on trouve 8 mots : accent, glissando, mélodie ou tunes, phonème tonal, registre, tessiture et ton. Par ailleurs, en A.P.I., il existe une extension des symboles et des signes diacritiques pour permettre la transcription étroite des énoncés. Parmi ces symboles, on retrouve certaines notations utilisées pour la transcription musicale et notamment, la transcription du tempo et des nuances. Ainsi, les variations de débit de parole sont indiquées par des indications de tempo comme **lento** (débit lent), **allegro** (débit rapide), **accelerando** (accélééré) et **rallentando** (ralenti) :

[XXX { <i>allegro</i> XXX <i>allegro</i> } XXX]	débit rapide
[XXX { <i>lento</i> XXX <i>lento</i> } XXX]	débit lent

Figure n°28 : indications de tempo utilisées pour la transcription du débit de parole en API.

Les variations d’intensité sont indiquées par des symboles de nuances comme **f** (**forte** : parole forte), **ff** (**fortissimo** : parole plus forte) et **p** (**piano** : parole douce) :

[XXX { <i>f</i> XXX <i>f</i> } XXX]	parole forte
[XXX { <i>ff</i> XXX <i>ff</i> } XXX]	parole plus forte
[XXX { <i>p</i> XXX <i>p</i> } XXX]	parole douce

Figure n°29 : indications de nuances utilisées pour la transcription des variations d’intensité dans la parole en API.

2.1.2. D’autres systèmes de transcription : leur validité par rapport à la transcription musicale

D’autres systèmes que la notation musicale ont été mis au point pour transcrire la prosodie. Nous n’en ferons pas le détail ici, mais on peut toutefois dégager trois grands types de transcription prosodique :

- On peut symboliser les mouvements mélodiques par des flèches, telles que ↑ pour indiquer une montée et ↓ pour indiquer une descente. Le système INTSINT⁹ (Hirst et Di Cristo, 1998) permet par exemple de transcrire l'évolution de la Fo sur une ligne séparée, sous la transcription orthographique ou phonétique (de manière à être indépendante de la division en groupes accentuels). Les flèches indiquent les points cibles, c'est-à-dire les pics ou les vallées de la courbe, qui sont considérés comme des tons (jugés comme les moments les plus informatifs du point de vue de la perception et de la production). Les symboles < (« Upstep ») et > (« Downstep ») indiquent une descente ou une montée par paliers, par terrasses (figure n°30).

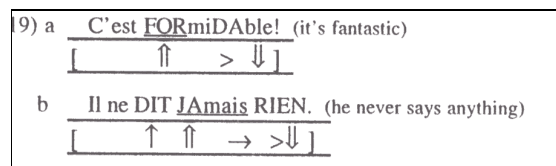


Figure n°30: transcription d'un énoncé à l'aide d'INTSINT (Di Cristo, 1998 : 209).

- A la place des flèches, on peut utiliser des lettres, comme **B** pour bas et **H** pour haut. C'est ce que fait par exemple le système ToBI¹⁰ (Silverman et al., 1992) pour l'anglais américain. Les points cibles mélodiques sont repérés par les lettres **H** (« High ») pour les tons hauts et **L** (« Low ») pour les tons bas. Les frontières sont indiquées par le symbole %.
- On peut tordre les lettres de manière à ce qu'elles représentent l'évolution de la courbe de Fo. C'est ce qu'a fait Bolinger (1998) pour décrire l'intonation de l'anglais américain (figure n°31). Ce faisant, il a réalisé un codage iconique de la Fo (Guaïtella, 2000 : 174), très facilement déchiffrable (cf. sur les portées primitives au IXe siècle, il n'y avait pas de neumes ou de notes, mais uniquement les syllabes du texte chanté, qui montaient ou qui descendaient en fonction de la progression de la notation musicale).

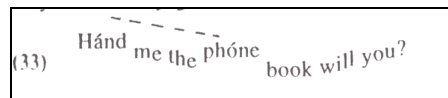


Figure n°31 : transcription de la mélodie par Bolinger (1998 : 51).

Tous ces systèmes de transcription ont été élaborés pour servir des visées théoriques bien précises. Mais est-ce que ces représentations se montrent aussi efficaces que la transcription musicale pour la lecture de la prosodie ?

⁹ « International Transcription System for Intonation ».

Pour le savoir, Fonagy et Bérard (1976) ont réalisé une étude pour comparer trois systèmes de transcription : la méthode graphique à cinq niveaux, les trois lettres H, M et B et la transcription musicale.

« Imaginons que les trois systèmes de transcription constituent trois canaux. Lequel des trois assure une meilleure transmission ? » (Guimbretière, 2000 : 157).

Pour le savoir, ils ont fait transcrire par des experts, dans chacun des trois systèmes, des énoncés assertifs, interrogatifs et répétitifs, des clichés mélodiques et des énoncés exprimant des attitudes spécifiques. Ils ont ensuite demandé à d'autres phonéticiens de reproduire la mélodie à partir de ces transcriptions. Les clichés mélodiques et les énoncés exprimant des attitudes spécifiques n'ont été correctement reproduits qu'à partir de la transcription musicale. Quant aux autres énoncés, la mélodie était aussi bien reproduite avec les trois systèmes. La transcription musicale s'avère donc le système le plus performant pour permettre l'oralisation du contour mélodique. Quant à la transcription par niveaux, on peut la considérer comme une variante de la transcription musicale, par rapport à des systèmes comme INTSINT et ToBI. La transcription musicale permet en outre une transcription « relativement » objective du contour mélodique. En effet, elle ne répond pas à une visée théorique spécifique de la prosodie comme c'est le cas pour INTSINT et ToBI. Avec ces systèmes de transcription, l'analyse est appliquée au moment même de la transcription. D'après Guaïtella (2000 : 173),

« L'intérêt d'une notation est de fournir le terrain le plus objectif possible pour permettre ensuite une interprétation plus libre. Si le système de codage se fonde sur des interprétations théoriques, alors l'analyse qui sera faite sur la base de ce codage présente une circularité dépendante du codage lui-même ».

¹⁰ « Tones and Break Indices ».

2.2. Lorsque la langue devient musique : Evolution du traitement de la similitude entre la langue et la musique au cours de l'histoire de la musique occidentale

« On pourrait se demander si la musique vocale ne représente pas un cas privilégié à l'intérieur de cette catégorie de systèmes, dans la mesure où elle unifie, en une temporalité unique, deux systèmes très différents. « Le langage est le plus vrai », disait Hegel, mais le langage isole, déplace, et à la limite je veux toujours autre chose que ce que je dis. « La musique exprime la pure vie intérieure », dit Boris de Schlæzer, mais elle est impuissante à nommer. On conçoit donc ce qu'il y a de séduisant dans une entreprise qui, en les combinant dans une fusion intime, grâce au truchement de leur organe commun, la voix, vise à donner l'illusion que, la béance qui est au cœur de l'un, l'autre viendra la combler, et réciproquement » (Ruwet, 1972 : 68).

Un compositeur pourra choisir de contraindre la ligne mélodique d'un texte parlé en accentuant la musicalité (fixation des intervalles et métricité du rythme) ou au contraire, la laisser évoluer plus « librement », comme elle le fait dans la parole naturelle. Cette dialectique constitue la trame de fond de l'histoire de la musique vocale occidentale, qui a toujours oscillé entre la priorité accordée à la mélodie (poème symphonique où le texte est absorbé par le niveau musical) et la priorité accordée à la prosodie (grégorien, récitatif), c'est-à-dire au texte. Pour Antonio Salieri, la musique doit primer sur la parole : « *primo la musica e poi le parole* » (1786). De même pour Mozart qui écrit que « *dans un opéra, il faut absolument que la poésie soit fille obéissante de la musique* » (lettre du 13 octobre 1781). Cependant, même chez Mozart, la musique doit savoir s'effacer au profit du texte dans le récitatif, de manière à ce que le spectateur puisse comprendre le déroulement de l'action. Chez un compositeur plus contemporain comme Debussy, l'écriture musicale est gouvernée par la musicalité de la langue. Il retrouve ainsi d'une certaine façon, près d'un millénaire après le chant grégorien, une musique du texte. Est-ce que la dialectique musique / texte tendrait à se résoudre dans la musique contemporaine ? C'est la question à laquelle nous essaierons de répondre dans les lignes qui vont suivre, mais quoiqu'il en soit, il est certain que son étude apporte un éclairage très subtil sur les rapports qu'entretiennent la langue et la musique.

2.2.1. Rapports texte et mélodie : information vs musicalité

2.2.1.1. Le chant grégorien, une musique du texte

Dans le chant grégorien, le texte revêt une importance fondamentale. Il s'agit en effet de transmettre les textes de la Bible aux fidèles par la lecture à haute voix et pour cela, l'intelligibilité des paroles doit être respectée. Cette exigence d'intelligibilité donne à la langue une place centrale. Le lecteur doit en connaître parfaitement les caractéristiques prosodiques s'il veut être compris par son public ; il doit prendre en compte la musique de la langue,

« ce chant latent (cantus obscurior) dont parle Cicéron, ou ce germe de musique (seminarium musices) selon l'expression de Martianus Capella » (Viret, 1987 : 38).

Bien qu'il subsiste quelques textes en grec, comme le « Kyrie Eleison », le latin d'église est la langue véhiculaire de cette liturgie. En conséquence, la musique appliquée à ces textes correspond à la musique de la langue latine, et notamment à son rythme :

« si ce texte influe sur la composition grégorienne au point de constituer son support essentiel, c'est qu'il ne se présente pas comme un simple assemblage de phonèmes que le compositeur traiterait à sa guise, mais bien comme un donné déjà investi par lui-même de caractères musicaux, principalement sous l'angle de la rythmique » (Viret, 1987 : 37).

Ce rythme d'essence linguistique évolue avec une certaine liberté et ne rentre pas dans un cadre métrique (cf. parole et périodicité sommaire, partie 1.2.1.3.). Ce cadre métrique est d'autant moins nécessaire que le chant grégorien est monodique, c'est-à-dire qu'il est chanté à cappella et sans accompagnement ; c'est un **plain-chant**. Il n'a donc pas besoin d'un rythme stable pour permettre de coordonner plusieurs voix entre elles, comme l'exige la polyphonie. Dans la langue, le rythme naît de la succession de syllabes accentuées et non accentuées (cf. définition du rythme, partie 1.2.1.). La rythmique du latin est organisée selon le retour des accents toniques qui affectent les mots. Pour cette raison, l'accent tonique est toujours indiqué dans les livres de chant.

Comment ces accents sont-ils traités dans le chant grégorien ? Viret (1987 : 39-40) nous donne un exemple du traitement des mots de 5 ou 6 syllabes. Ces mots peuvent être accentués soit sur l'antépénultième (mots proparoxytons), soit sur la pénultième (mots paroxytons). Les deux ou trois syllabes initiales préparent l'arrivée de l'accent, point culminant du mot.

Ce temps d'attente créé chez l'auditeur un sentiment de tension, qui ne peut s'apaiser qu'au moment où l'accent survient. Les syllabes atones qui suivent viennent conclure le mouvement et concourent ainsi à la détente survenue après l'arrivée de l'accent. L'accent constitue donc le « pôle rythmique » du mot et le schéma mélodique lui est subordonné : l'intonation monte avant l'accent, atteint son pic sur l'accent, puis redescend sur les syllabes atones terminales (figure n°32).

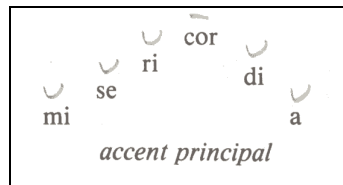


Figure n°32: schéma accentuel du mot « misericordia » (emprunté à Viret, 1987 : 40).

L'adjonction d'un accent secondaire sur la syllabe initiale peut venir complexifier la simplicité de ce schéma, par un dédoublement de la courbe (figure n°33), mais le pattern mélodique garde cependant sa configuration « en cloche ». Ces schémas de montée et de descente reposent sur les mêmes principes psychologiques que les schémas de tension et de détente dans la perception de la mélodie (cf. partie n°1.2.2).

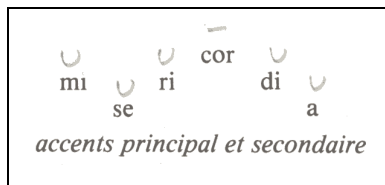


Figure n°33 : schéma accentuel du mot « misericordia » (tiré de Viret, 1987 : 40).

Selon Fraisse (1974 : 134),

« nul isochronisme n'apparaît. L'important reste un mouvement intensif et mélodique. Dom de Malherbe évoque, pour décrire ce rythme libre, les mouvements de la nature. « Il faut chanter comme une vague bondit ou s'étale, comme un champ de blé ondule, comme un arbre s'agite sous l'effort du vent, comme un fleuve roule ses eaux » (1946 : 24). Mais tous les restaurateurs du chant grégorien s'accordent pour invoquer le vol d'un oiseau comme la meilleure métaphore : après les syllabes préliminaires, la voix monte vers la syllabe accentuée, comme dans un élan, puis redescend vers le repos où se produit une nouvelle inspiration ».

La notation neumatique utilisée pour la transcription du chant grégorien ne comporte aucune indication de rythme, tout simplement parce que ce rythme est celui de la langue. Il suffit donc au chanteur de bien connaître la rythmique du latin.

Si le rythme du chant grégorien reste avant tout celui du mot latin, c'est sa mélodie qui va le faire évoluer vers une « musicalité » grandissante. Plusieurs facteurs concourent à cette évolution. Le premier d'entre eux concerne l'évolution du style du chant. Au départ, les chantres récitaient le texte sur une seule note, ce qui devait être relativement monotone pour les auditeurs. Mais, peu à peu, ils vont augmenter le ton aux endroits où la mélodie monte naturellement et le baisser, là où elle descend naturellement. Ces montées et ces descentes correspondent à des frontières linguistiques importantes. Comme pour le rythme, la ligne mélodique est donc calquée sur l'intonation de la langue latine. Le chant est syllabique (**extrait musical n°5** et partition correspondante, figure n°34), dans la mesure où on chante une note par syllabe, comme c'est le cas dans les antiennes (le chant qui sert d'introduction et de conclusion à la psalmodie, chant des psaumes).

Figure n°34 : chant syllabique - Sanctus XIII, VIIIe Mode (tiré du livret d'accompagnement du CD « Chant Grégorien, Chant de Toujours », p : 15).

Mais peu à peu, la mélodie va se complexifier. Le nombre de notes par syllabes va augmenter pour donner naissance aux ornements, les **mélismes**. La figure n°35 nous en donne un exemple : sur la voyelle « e » de « Kyrie », la ligne mélodique se prolonge sur une douzaine de notes (**extrait musical n°6** et partition correspondante, figure n°35).

Figure n°35 : chant mélismatique - Kyrie XIII, IVe Mode (tiré du livret d'accompagnement du CD « Chant Grégorien, Chant de Toujours », p : 13).

Ce chant mélismatique nécessite une technique vocale beaucoup plus poussée que dans le chant syllabique, ce qui va transformer les moines lecteurs en véritables moines chanteurs. Les mélismes préfigurent déjà les **vocalises** de l'opéra italien. On les retrouve dans l'alléluia (notamment dans le « jubilus »), qui est un chant d'acclamation de louanges.

Mais, cependant, il y a une limite à cette ornementation : le chant grégorien doit resté compréhensible, car il est avant toute chose, une musique du texte. La compréhensibilité constitue donc un frein naturel à toute évolution plus radicale du chant grégorien.

« Rien de tout cela ne saurait éclipser le sens du texte liturgique revêtu d'une mélodie destinée à le commenter, l'éclairer, le mettre en valeur. C'est donc lui qu'il faut interroger d'abord... On ne saurait être trop catégorique sur ce point » (Dom Gajard, 1951 : 39¹¹).

Un autre facteur contribue à faire évoluer la mélodie et il est d'une importance capitale dans l'histoire de la musique occidentale : au début du chant grégorien, les hauteurs ne sont pas encore fixées par la notation musicale (cf. partie 1.2.2.2.). L'apparition d'une hauteur référence va permettre cette fixation, et par conséquence, la fixation des intervalles. Cette évolution de l'écriture pousse inéluctablement la mélodie vers une plus grande « musicalité » (cf. 1.2.4.) et concourt à creuser le fossé entre la langue et la musique.

2.2.1.2. Le récitatif dans l'opéra

2.2.1.2.1. A la frontière de la parole et de la musique

Progressivement, la musique vocale évolue vers la **polyphonie** et cette multiplication des voix va venir brouiller l'intelligibilité du texte. La polyphonie nécessite une rythmique plus métrique nécessaire pour que les différentes voix s'accordent entre elles, mais très éloignée de la liberté rythmique de la parole. Ces bouleversements éloignent de plus en plus la musique du texte, ce qu'illustrent les oeuvres de l'Ecole Franco-Flamande et notamment, celles de Roland de Lassus (1531-1594) et de Giovanni Palestrina (1525-1594). Chez ces compositeurs, l'équilibre de l'architecture polyphonique prime sur la compréhension des paroles et l'interprète perd son individualité. En fait, il n'a d'existence qu'au sein du groupe vocal. C'est en réaction contre cette complexification de l'écriture vocale qu'apparaît le **récitatif**.

A Florence, un cercle d'humanistes rassemblés autour du Comte Giovanni Bardi (Camerata Bardi) entreprend de rénover la musique vocale selon le modèle de la tragédie grecque classique. Selon eux, il faut retrouver l'intelligibilité du texte en revenant à la monodie :

¹¹ Le chant grégorien a été rénové au cours du XXe siècle principalement par les moines de l'Abbaye de Solesmes. La théorie de Dom Mocquereau sur le rythme libre musical s'illustre particulièrement et a été sa mise en musique par le maître de chœur de Solesmes, Dom Gajard dans les années 60.

« *il faut renoncer au contrepoint et revenir à la simplicité du mot* » (Vincenzo Galilei, Camerata Fiorentina).

Autrement dit, la musique doit se plier au texte, elle doit servir la poésie.

Jacopo Peri qui fréquente la Camerata Bardi sera le premier compositeur à appliquer ces préceptes dans les deux premiers opéras de l'histoire de la musique, « *Dafne* » (1597) et « *Euridice* » (1600). « *Euridice* » se caractérise par l'utilisation d'un nouveau style, le *recitar cantando* qu'on pourrait traduire par « *parler en chantant* ». Ce nom illustre parfaitement l'ambiguïté d'un style qui neutralise les oppositions habituelles entre la lecture et le chant. Non seulement, la musique se lie au verbe, mais elle devient celle du verbe : la ligne mélodique et l'écriture rythmique suivent fidèlement l'intonation et le rythme de la langue utilisée dans le texte. Comme dans le chant grégorien, l'interprète doit connaître les caractéristiques prosodiques de la langue qu'il emploie. En se transformant en orateur-chanteur, il recouvre la liberté qui est propre à la parole : c'est lui qui imprime le rythme, qui « conduit » l'accompagnement. Il n'y a plus besoin du chef d'orchestre car les instruments accompagnateurs s'alignent sur le chanteur. Le texte redevient central et avec lui, l'information qu'il véhicule. Un des problèmes majeurs dans « *Euridice* » vient justement du fait qu'il ait été presque entièrement écrit en *stile recitativo*. En effet, il souffre d'un manque de musicalité et d'expressivité.

Ce déséquilibre sera rectifié dans l'« *Orfeo* » (1607), le premier chef d'œuvre de l'opéra. Son auteur, Claudio Monteverdi n'utilise plus systématiquement le *recitar cantando*, mais il a également recours au contrepoint pour les parties symphoniques et au **madrigal** pour les parties chorales. L'**aria** n'apparaît pas encore, mais peu à peu, l'expression des sentiments va lui être confiée. La musicalité y reprend le dessus souvent aux dépens de la compréhension du texte. Ce tiraillement entre l'exigence d'intelligibilité nécessaire à la compréhension du drame et la musicalité structure toute l'évolution de l'opéra italien au cours du XVIIe siècle. Cette lutte ne prendra fin que lorsque la séparation de l'air et du récitatif sera consommée, c'est-à-dire à partir des années 1650-1660, et notamment dans l'œuvre d'Antonio Cesti (1623-1669). A partir de cette époque, l'aria et le récitatif ont des fonctions nettement délimitées dans l'opéra.

« *Dans l'opéra italien, la fonction du récitatif est de relier les airs en exposant les péripéties de l'action sous forme de monologue ou de dialogue* » (Massin et Massin, 1983 : 98).

Il doit donc être le plus intelligible possible, afin que les spectateurs puissent suivre le cours du récit. L'aria quant à lui permet au chanteur de donner libre cours à sa virtuosité et de montrer sa maîtrise de la technique vocale. Il constitue selon Massin et Massin (1983 : 97) « *la friandise des gourmets de la musique lyrique* ».

En France, c'est Jean-Baptiste Lully qui fait apparaître le récitatif dans la tragédie lyrique. Ce récitatif qu'on appelle « français » suit également le texte, mais se montre plus mélodique que son homologue italien. Au XVIII^e siècle, on trouve deux types différents de récitatifs : le **recitativo secco** qui est un récitatif simple, où la voix est accompagnée d'un clavecin ou d'une basse continue et le **recitativo accompagnato**. Ce dernier type de récitatif est accompagné par tout l'orchestre. Son rythme est moins libre que le recitativo secco, mais il se montre plus expressif et plus musical. Il se situe à l'intermédiaire du recitativo secco et de l'aria. Dans l'opéra bouffe dont le « *Don Giovanni* » de Mozart marque le sommet, le récitatif adopte la forme « *parlando* ». Son écriture est tellement souple et l'accompagnement tellement réduit (le plus souvent limité au clavecin) qu'il permet de restituer la rapidité du dialogue naturel, surtout dans les dialogues entre les personnages (**extrait musical n°11**, dialogue de Leporello et Don Giovanni, acte II, scène XI).

2.2.1.2.2. Trois récitatifs, trois langues, trois musiques

Comme le chant grégorien, le récitatif est une musique du texte qui montre une proximité maximale avec la parole. Certains types de récitatifs sont plus musicaux que d'autres, comme le recitativo accompagnato, mais ils se situent tous dans une zone frontière où l'on ne sait plus très bien distinguer la langue de la musique. Dans ce contexte, le problème de la musicalité intrinsèque des langues se pose avec une grande force. Il est évident que l'allemand ne chante pas comme l'anglais, qui ne chante pas comme l'italien. Les règles d'accentuation et les contraintes rythmiques sont différentes et chacune de ces trois langues possède sa prosodie spécifique. Comment vont-elles être « mises en musique » dans le récitatif ? Est-ce que leur musique intrinsèque va être respectée par le compositeur ? Pour le savoir, nous avons choisi trois récitatifs écrits dans trois langues différentes : ils sont tirés des œuvres « *Didon et Enée* » de Henry Purcell en anglais (**extrait sonore n°7**), les « *Passions de St Matthieu* » de Jean-Sébastien Bach en allemand (**extrait sonore n°9**) et « *Don Giovanni* » de Wolfgang Amadeus Mozart en italien (**extrait sonore n°11**). Le texte de chacun de ces récitatifs a ensuite été lu par trois locuteurs natifs anglais (**extrait sonore n°8**), allemand (**extrait sonore n°10**) et italien (**extrait sonore n°12**). Ces enregistrements nous permettront de comparer l'accentuation et l'intonation adoptées par le chanteur avec celles adoptées par le lecteur.

2.2.1.2.2.1. Anglais : « *Didon et Énée* » de Purcell

Didon et Enée est un opéra en trois actes créé par le compositeur anglais Henry Purcell à Chelsea, en 1682. Dans cette œuvre, Purcell utilise une forme de récitatif dont la ligne mélodique évolue plus souplesment que dans le recitativo secco. L'extrait intitulé « *Ruin'd ere the set of Sun ?* » est tiré de la scène de la grotte, qui prend place dans le deuxième acte de l'opéra et met en scène des sorcières et des enchanteurs. Le passage choisi concerne une des répliques des sorcières (**extrait sonore n°7** et figure n°36). La ligne mélodique est chantée, mais elle suit très fidèlement l'intonation de la langue (pour une description en détail du système prosodique de l'anglais, voir chapitre V). Ainsi, dans « *Trustly elf* », la ligne mélodique s'infléchit sur la syllabe atone « *ly* », comme elle le fait dans la langue (**extrait sonore n°8**). De même, la tendance de l'anglais à accentuer la syllabe initiale dans les mots lexicaux plurisyllabiques est respectée (« *MERcury* », « *TRUSTly* »). Quant à sa rythmicité accentuelle, elle apparaît moins au niveau de la partition, qu'au niveau de l'interprétation (figure n°37). Il est vrai que certaines syllabes accentuées sont allongées, comme « *elf* », « *self* », « *chide* » ou « *night* ». La valeur des notes correspondantes équivaut à une noire (un temps) au minimum, alors que pour les autres syllabes, les valeurs sont plus courtes, c'est-à-dire des croches (un demi-temps) ou des doubles-croches (un quart de temps). Cependant, les autres syllabes accentuées ne se démarquent pas sur la partition. C'est le chanteur qui les met en relief en les produisant avec une plus grande force articulatoire, comme dans « *As SENT from JOVE* ». Tous les mots lexicaux sont accentués et les syllabes atones sont réduites comme la syllabe « *ly* » de « *trustly* », ce qui n'apparaît pas non plus sur la partition. L'accompagnement instrumental suit l'accentuation linguistique et le rythme imprimé par le chanteur : les envolées du clavecin commencent toujours sur une syllabe accentuée ; quant à la ligne mélodique de la basse et des cordes, elle ne change que sur des syllabes accentuées. On observe un changement de rythme dans l'accompagnement à partir de « *sail tonight* » avec l'introduction d'un accent d'insistance sur « *all* » exprimé musicalement par un pic mélodique (montée d'une quarte pour aboutir au ré3 correspondant à la syllabe « *all* », puis descente d'une quinte, pour aboutir au sol3) et un allongement (croche pointée), suivie d'une imitation musicale des rires des sorcières. Dans tout l'extrait, il y a très peu d'ornementation, sauf sur les deux syllabes « *chide* » (appoggiature et trille) et sur « *charge* » (notes de passage). Ces deux syllabes étant « naturellement » allongées parce qu'elles portent l'accent, elles laissent le temps au chanteur de charger un peu la ligne mélodique. Chacune des autres syllabes de l'extrait correspond à une seule note, ce qui rend le texte très intelligible.

Si l'on compare cet extrait au texte lu par le locuteur anglophone (**extrait sonore n°8**), on relève une correspondance quasi parfaite dans la localisation des accents (voir tableau n°31). Tous les pics mélodiques correspondant aux syllabes accentuées dans le texte parlé, n'apparaissent pas dans le texte chanté. Cependant, le chanteur réalise les accents de mot en allongeant les syllabes.

Figure n°36 : partition de l'extrait, « Ruin'd ere the set of Sun ? », tiré de Didon et Aenée de Purcell (1682 : 32-33).

MUSIQUE (extrait sonore n°7)	PAROLE (extrait sonore n°8)
<p>Sorceress¹² (extrait n°7) :</p> <p>But when they've done ↑, my trustly elf, In form of Mercury himself ↑, As sent from Jove, shall chide ↑↑ his stay, And charge him sail tonight ↑ With all ↑ this fleet away.</p>	<p>Sorceress (extrait n°8) :</p> <p>But when ↑ they've done ↑, my trustly elf, ↑ In form ↑ of Mer↑cury himself ↑, As sent ↑ from Jove↑, shall chide ↑ his stay↑, And charge ↑ him sail ↑tonight ↑ With all ↑ this fleet away.</p>

Figure n°37 : accentuation comparée des extraits chantés et parlés de « Ruin'd ere the set of Sun ? », les syllabes accentuées étant indiquées en caractères gras et les pics mélodiques par une flèche montante.

¹² **Traduction** : Mais, dès qu'ils en auront fini, mon fidèle lutin – Sous la forme de Mercure lui-même – Envoyé par Jupiter, lui reprochera de rester – Et lui ordonnera de faire voile ce soir avec toute sa flotte.

2.2.1.2.2 Allemand : « *Les Passions de St Matthieu* » de Bach

Le récitatif suivant est intitulé « *Da ging him der Zwölfen einer* » (**extrait sonore n°9**). Il est tiré des « *Passions de Saint Matthieu* » (appelée également Grande Passion), composées en 1727 par Jean-Sébastien Bach. L'extrait met en scène l'évangéliste et l'un des disciples de Jésus, Judas. L'évangéliste raconte que Judas est venu trouver les prêtres pour leur proposer de leur livrer Jésus, en échange de 30 pièces d'argent. La narration est interrompue par la question de Judas aux prêtres en style direct : « *Was wollt ihr mir geben ? Ich will ihn euch verraten* ». La ligne mélodique de l'Évangéliste est « chantée », mais l'intervention de Judas est très proche du langage parlé, surtout au début de la question : « *Was wollt ihr mir geben ?* », en recitativo secco. Par rapport au récitatif de Purcell, l'accompagnement est beaucoup plus sommaire. Dans la première intervention de l'évangéliste correspondant à la phrase « *Da ging hin der Zwölfen einer, mit Namen Judas Ischarioth, zu den Hohen Priestern und sprach* », les accords sont plaqués par la basse continue (ici, les cordes) et par l'orgue sur les syllabes qui en délimitent le début et la fin, c'est-à-dire « *Da* » et « *sprach* ». Quant à la syllabe « *rioth* », elle indique une frontière importante correspondant à la fin de la première proposition (voir partition, figure n°38).

11* [RECITATIVO]

EVANGELIST

Then went one of the di-sci-ples, whose name was Ju-das Is-ca-ri-ot, to the chief. priests and—
 Chorus I Da ging hin der Zwöl-fen ei-ner, mit Na-men Ju-das Is-cha-ri-oth, zu den Ho-hen prie-ster-n, und

4 JUDAS EVANGELIST
 said, How much will ye give me if I de-liv-er Him to you? And they of-fered him thir-ty
 sprach: Was wollt ihr mir ge-ben? Ich will ihn euch ver-ra-ten. Und sie bo-ten ihm drei-ßig

7 sil-ver pie-ces. And from that time he sought op-por-tun-i-ty that he might be-tray Him.
 Sil-ber-lin-ge. Und von dem an such-te er Ge-le-gen-heit, daß er ihn ver-rie-te.

Figure n°38 : partition de l'extrait, « Da ging hin der Zwölfen einer », tiré des « Passions de Saint Matthieu » (1727, Bach, p : 36).

A partir de l'intervention de Judas, les accords se font plus fréquents et correspondent aux syllabes accentuées « will », « ra » de « verraten », « bo » de « boten » et « drei » de « dreissig ». Cette multiplication correspond à un rythme plus saccadé, plus proche de la parole naturelle. L'accompagnement colle donc totalement à la structure accentuelle du texte et ponctue le mouvement rythmique imprimé par les chanteurs. Par ailleurs, on observe une correspondance frappante (voir figure n°39) entre les syllabes accentuées de l'extrait chanté (**extrait sonore n°9**) et celles de l'extrait parlé par la locutrice allemande (**extrait sonore n°10**). Sur la partition, certaines syllabes atones, comme dans la parole, sont très compressées au niveau temporel (durée d'une double croche, c'est-à-dire un quart de temps), comme c'est le cas des syllabes « ihr » et « mir » dans « Was wollt ihr mir geben ? ». La rythmique est quasiment identique, c'est la rythmique spécifique de l'allemand. En ce qui concerne l'intonation, les pics mélodiques de l'énoncé chanté « Was wollt ↑ ihr mir ge-↑ben ? » sont également en parfaite correspondance avec l'énoncé parlé. Ce passage est justement en discours direct, comme si Bach avait voulu qu'il soit le plus proche possible de la parole. En revanche, dans le reste du texte, on observe des différences de localisation des pics

mélodiques avec le parlé : ils se répartissent à des endroits différents. Au niveau de l'interprétation, les deux chanteurs ne se permettent aucune fioriture, juste une petite note de passage sur « *riot* ». Ils s'en tiennent strictement à une note par syllabe ce qui donne au texte une intelligibilité maximale.

MUSIQUE	PAROLE
Evangelist : Da ging hin der Zwöl-↑ fen einer, mit Namen Judas Ischari-↑ oth, zu den Hohenprie-↑ stern, und sprach ↑↑ . Judas : Was wollt ↑ ihr mir ge-↑ ben ? Ich will ihn euch ↑↑ verraten. Evangelist : Und ↑ sie boten ihm drei-↑ ßig Silber-↑ linge. Und ↑↑ von dem an suchte er Gele-↑ genheit, daß er ihn verrie-↑ te.	Evangelist : Da ging ↑↑ hin der Zwölfen einer, mit Namen Ju-↑das Ischarioth, zu den Hohenpriestern ↑, und sprach . Judas : Was wollt ↑ ihr mir ge-↑ben ? Ich will ↑ ihn euch verraten. Evangelist : Und sie bo-↑↑ ten ihm drei-↑ ßig Silberlinge. Und von dem ↑↑ an suchte er Gele-↑ genheit, daß er ihn ver-↑ riete.

Figure n°39 : accentuation comparée des extraits chantés et parlés de « *Da ging him der Zwölfen einer* », les syllabes accentuées étant indiquées en caractères gras et les pics mélodiques par une flèche montante.

2.2.1.2.2.3. Italien : « *Don Giovanni* » de Mozart

Le dernier extrait est tiré du célèbre opéra composé par Mozart « *Don Giovanni* » (Prague, 1787). C'est un opéra en deux actes, dont le livret a été écrit par Lorenzo da Ponte. L'extrait intitulé « *Ah ah ah ! ah, questa è buona* » prend place dans le deuxième acte, scène XI. Il relate une conversation entre Don Giovanni et son valet, Leporello (**extrait sonore n°11**). Le récitatif employé par Mozart est un Recitativo secco sous la forme parlando. Des trois récitatifs présentés, c'est celui qui restitue le mieux le naturel de la parole. La situation dans laquelle prend place ce dialogue exige une rapidité d'échange entre les deux personnages et une grande spontanéité dans l'expression des émotions (surprise, colère). Le spectateur doit « y croire » et c'est ce que Mozart s'est employé à faire. Le résultat est très réaliste. Le début du récitatif commence par l'interpellation de Leporello par Don Giovanni : « *Oh, Leporello !* », tout en douceur (la scène se passe de nuit et Don Giovanni veut se montrer très discret). La réaction de Leporello exprime la surprise « *Chi mi chiama ?* ». Ces deux phrases sont du parlé pur. La suite consiste en un très subtil dosage entre la déclamation et l'introduction d'un peu de mélodicité dans le phrasé. Dans la version sonore, l'accompagnement est quasiment absent et laisse une totale liberté aux deux interprètes. En effet, le violoncelle et la contrebasse plaquent des accords tous les 8 temps. Sur la partition (figure n°40), ces accords sont tenus pendant 8 temps (deux rondes liées), ce qui est un peu plus « lourd ».

Figure n°40 : partition de l'extrait, « Ah ah ah !ah, questa è buona », tiré de Don Giovanni de Mozart (1787 : 297)

Si on compare maintenant l'accentuation de l'extrait musical (**extrait sonore n°10**) à celle de l'extrait produit en version parlée par une locutrice italienne native (**extrait sonore n°11**), on remarque une très forte correspondance (figure n°41), comme dans le récitatif précédent.

MUSIQUE	PAROLE
Don Giovanni : Oh Leporello ! ¹³	Don Giovanni : Oh Leporello !
Leporello : Chi mi chiama ?	Leporello : Chi mi chiama ?
Don Giovanni : Non conosci il padron ?	Don Giovanni : Non conosci il padron ?
Leporello : Così nol conoscesi !	Leporello : Così nol conoscesi !
Don Giovanni : Come ? birbo !	Don Giovanni : Come ? birbo !
Leporello : Ah siete voi ? Scusate !	Leporello : Ah siete voi ? Scusate !
Don Giovanni : Cosa è stato ?	Don Giovanni : Cosa è stato ?
Leporello : Per cagion vostra io fui quasi accoppato.	Leporello : Per cagion vostra io fui quasi accoppato.
Don Giovanni : Ebben , non era questo , un onore per te ?	Don Giovanni : Ebben , non era questo , un onore per te ?
Leporello : Signor , vel dono !	Leporello : Signor , vel dono !
Don Giovanni : Via , via , vien qua , vien qua : che belle cose ti deggio dir !	Don Giovanni : Via , via , vien qua : che belle cose ti deggio dir !

Figure n°41 : accentuation comparée des extraits chantés et parlés de « Ah ah ah !ah, questa è buona » les syllabes accentuées étant indiquées en caractères gras et les pics mélodiques par une flèche montante.

¹³ Traduction en français : Hola, Leporello ! – Qui m'appelle ? – Tu ne connais pas ton maître ? – Poussais-je ne pas le connaître ! – Comment ? Coquin - Ah, c'est vous ? Excusez ! – Qu'y a-t-il eu ? – A cause de vous, j'ai presque été assommé – Eh, bien, ce n'était pas là un honneur pour toi ? – Monsieur, je vous en fait cadeau ! – Allons, allons, viens ici : que de belles choses j'ai à te dire !

Dans l'extrait chanté, l'accentuation de l'italien est donc respectée par l'interprète, mais elle apparaît d'une manière beaucoup plus marquée dans la notation musicale de la partition qu'elle ne l'était dans le récitatif de Bach. En italien, la syllabe pénultième des mots plurisyllabiques est accentuée et notamment marquée par un allongement. Dans la notation, les syllabes accentuées sont presque systématiquement allongées (croches, voir noires), alors que les syllabes atones sont réalisées sur des double-croches, là-aussi quasiment systématiquement. C'est cette alternance entre valeurs longues et valeurs courtes qui permet à la notation musicale de restituer si fidèlement le rythme de l'italien parlé (forme parlando). Dans le récitatif de Bach, on peut dire que le chanteur ajoute sa propre connaissance de la prosodie de l'allemand à l'interprétation, il doit donc connaître l'accentuation de l'allemand. Dans le récitatif de Mozart, l'interprète n'a à la limite pas besoin de connaître l'accentuation de l'italien puisque celle-ci est restituée presque parfaitement par la notation musicale.

La rythmicité et le système intonatif de chacune de ces langues ont été respectés lorsque le compositeur les a mis en musique dans le récitatif, ce qui assure au texte une intelligibilité maximale. Les différences prosodiques entre ces langues engendrent donc des mises en musique différentes. Des trois récitatifs, celui de Mozart restitue le mieux la parole et ses changements de débit selon les émotions qui affectent les personnages (rapidité d'élocution sous le coup de la colère par exemple). La notation musicale se trouve donc ici appropriée à la restitution de la prosodie et de l'effet « parlando ». Ce n'est plus le texte qui est mis en musique, mais la musique qui est mise en texte. On trouve un seul exemple dans la musique instrumentale « classique », qui ait été traité par le compositeur en **mode récitatif** pour imiter la parole naturelle : c'est le début de l'« *Ode à la Joie* » de Schiller, dans la « *Neuvième Symphonie* » de Beethoven où le violoncelle et la contrebasse préparent l'entrée du chanteur baryton en mode de récitatif (**extrait sonore n°13**).

2.2.1.3. La musicalité des langues dans la mise en musique

2.2.1.3.1. Langues musicales versus langues non musicales ?

Le chant grégorien est fondé sur le rythme et l'intonation de la langue latine. Il en est de même dans les trois récitatifs précédents dont la ligne mélodique et le rythme suivent le « chant » de l'anglais, de l'allemand et de l'italien. Il est donc impossible de traduire ces textes dans une autre langue sans en modifier la musique.

Le problème de la musicalité des langues et de la gestion de cette musicalité dans la composition est central dans la musique vocale. Est-ce que certaines langues se plient plus facilement à la mise en musique ? Est-ce qu'on peut passer d'une langue à une autre sans modifier la musique, ou au contraire, faut-il l'adapter à la nouvelle langue ? Ces questions semblent s'être souvent posées aux compositeurs, au cours des siècles.

La **Querelle des Bouffons** aborde ce problème de manière caricaturale. Elle débuta en 1752 et opposa les partisans de l'opéra italien (ou « *opera buffa* ») à ceux de l'opéra français. Jean-Jacques Rousseau, grand amateur d'opéra italien s'illustra dans cette querelle par l'ensemble des critiques qu'il porta à l'opéra français dans sa « *Lettre sur la Musique Française* » (1753), et notamment contre la musique de Jean-Philippe Rameau (1683-1764). L'une de ses critiques porte sérieusement à discussion. Il reprochait notamment à la langue française de ne pas être une langue musicale, contrairement à l'italien. Compositeur lui-même, il mis sa théorie en pratique dans « *Pygmalion* » (1770), œuvre composée en français, mais où « *les paroles et la musique, au lieu de marcher ensemble, se font entendre successivement* ».

Ce faisant, Rousseau soulève le délicat problème de la musicalité des langues. Il est vrai que chaque langue possède sa propre musique, mais de là à juger que certaines langues en soient totalement dépourvues, il y a un pas qu'un linguiste ne peut franchir. Il n'existe pas de langue « non musicale » ; en revanche, toutes les langues ont en commun de combiner différemment les trois paramètres que sont la hauteur, la durée et l'intensité. L'organisation prosodique qui résulte de cette combinaison peut nous paraître plus ou moins musicale. Par exemple, l'anglais se caractérise par de larges variations de hauteur et une accentuation plus marquée du point de vue acoustique que le français (voir chapitre V, partie 2.1.1.2), ce qui pousse un compositeur comme Claudy Malherbe à adopter cette langue pour la composition de son œuvre « *Locus* » (voir plus bas, partie 2.2.3.3.). Le français serait-elle donc une langue moins musicale que l'anglais ? C'est ce que sous-entend Stricker lorsqu'il reproche à Maurice Ravel de s'être appuyé sur la prosodie du français dans la composition de ses œuvres. Il met notamment en avant le problème de la faiblesse de l'accent tonique :

« Ravel s'est employé à calquer sa ligne mélodique sur celle de la phrase parlée, s'inspirant des faibles accentuations toniques de la langue française au point de subordonner l'ampleur de la ligne mélodique à cet impératif du texte. Là peut-être pourrait-on voir une limite trop définie au génie mélodique ravélien, car cette faiblesse de l'accent tonique qui caractérise le français bride toujours quelque peu une ligne vocale qui s'en inspirerait, en même temps qu'elle défie par là une véritable transposition musicale ».

Mais, on peut se demander à l'inverse si ce « génie ravélien » dont parle Stricker aurait pu exister dans une autre langue que le français ? Prenons une œuvre comme « *Pelléas et Mélisande* » de Claude Debussy (voir plus bas, partie 2.2.1.2.). Il est évident qu'une telle composition puise ses qualités musicales dans celles de la langue française. On peut même dire que la grande valeur artistique de cette œuvre résulte de la fusion parfaite entre la langue et la musique et « *Pelléas et Mélisande* » ne saurait exister en anglais. Non, décidément, les rapports entre le texte et la musique paraissent infiniment plus subtils. En fait, il semble que dans la musique vocale, le problème ne se réduise pas à déterminer si certaines langues sont plus musicales que d'autres, mais plutôt à savoir si le mariage entre paroles et musique est plus ou moins bien assorti. Pour arriver à une cohésion entre les deux systèmes, le compositeur doit prendre en compte les spécificités prosodiques de la langue dans laquelle il compose, car l'accentuation musicale ne doit pas entrer en contradiction avec l'accent de mot ou l'accent de phrase. Mais présenté comme cela, la prosodie paraît ne représenter qu'une source de contrainte pour le compositeur. Il faut souligner qu'elle lui apporte également une grande source d'inspiration pour l'écriture de la partie musicale. Pour mettre en harmonie la langue et la musique, outre ses qualités musicales, le compositeur doit donc aussi avoir l'âme d'un phonéticien.

2.2.1.3.2. Un obstacle à la traduction des œuvres musicales ?

Une œuvre vocale composée dans une langue spécifique n'est donc pas transposable dans une autre langue, puisque celle-ci chante différemment. Le compositeur est de ce fait obligé d'adapter son écriture musicale pour qu'elle soit en correspondance avec le nouveau texte ou alors, de retravailler le texte jusqu'au point où il puisse s'accorder avec la musique. Nous l'avons vu, traduire « *Pelléas et Mélisande* » en allemand ou en anglais implique de réécrire totalement la partition, ce qui exige un travail titanesque de la part du compositeur. C'est pourtant ce que fit Giuseppe Verdi (1813-1901) dans son « *Don Carlos* » (1867), en donnant deux versions de son œuvre, une en italien et l'autre en français.

Annie Prassoloff (mars 1999¹⁴) souligne le « *tour de force technique, une exploration du génie des deux langues, de leurs vertus et de leurs limites* » réalisé par le compositeur. Pour illustrer les différences entre les deux langues, elle prend pour exemple le grand air de Philippe II et le compare dans les deux versions :

« le français donne au grand air de Philippe II une sorte de nudité aggravée par la gaucherie de l'accentuation : "Elle ne m'aime pas...", alors que l'italien magnifie le lyrisme désolé de l'aveu en mettant au premier plan le maître mot et ses voyelles pleines : "Amor per me non ha" ».

Pour Verdi, la traduction du texte impose donc une mise en musique différente.

Pourtant, il se trouve des auteurs qui nient le problème de la musicalité des langues, donc de la traduction. Selon Boris de Schlœzer, les paroles se dissolvent complètement dans la musique :

« Le sens des paroles chantées n'est plus le même que celui qu'avaient ces mêmes paroles avant leur mise en musique... mais celui que leur confère la phrase musicale » (19 ??: 273).

Il apporte pour preuve que, dans l'histoire de la musique, on trouve des textes identiques mis en musique différemment et inversement, une même musique servie par des textes différents. Mais, si le choix du texte n'a aucune importance, alors « *pourquoi donc les musiciens ont-ils toujours mis tant de soin à choisir leurs textes et leurs livrets, voire à les composer eux-mêmes ?* » C'est la question que se pose Ruwet (1974 : 42-43) en commentant la position radicale de Schlœzer :

« En lisant Boris de Schlœzer, on a souvent l'impression qu'à écrire de la musique vocale, on ne risque que des inconvénients, des malentendus, des limitations. Pourquoi les affronter si on ne doit pas aussi y gagner quelque chose de plus fondamental ? ».

Même dans le récitatif secco sous sa forme parlando, la musique ne s'efface jamais complètement devant la langue et inversement dans le cas extrême de l'aria. Selon Ruwet (1972 : 52),

¹⁴ Texte non publié, disponible sur le web à l'adresse suivante : <http://www.chez.com/annieprassoloff/articles/texte-et-musique.htm>.

« dans la mesure où la voix est, pour l'homme, avant toute chose, l'organe de la parole, dès qu'elle intervient en musique, le langage, comme tel, est présent, et cela même si le chant s'émancipe en de purs mélismes, même si le texte devient totalement incompréhensible. [...] Il semble qu'il soit impossible de voir dans la voix un instrument comme les autres ».

2.2.2. La crise de la mélodie ou le retour à l'intonation

Grâce à son écriture (partie 1.2.2.2.), la musique a progressivement fixé les intervalles et par conséquent, donné naissance à l'harmonie, dont l'intonation mélodique est la forme la plus élémentaire. **Le développement de l'harmonie s'est fondé sur la hiérarchie tonale, c'est-à-dire sur la hiérarchie entre les 7 degrés de la gamme** (partie 1.2.2.4.). Mais, peu à peu, le système tonal classique va pouvoir s'élargir. Pour des raisons techniques (facture des instruments, mise au point du diapason), il n'a pas été possible d'utiliser toutes les tonalités avant le XVIII^e siècle. La mise au point de la gamme tempérée (division de l'octave en 12 demi-tons égaux) a permis au système tonal d'atteindre sa plénitude, mais l'a conduit également à la saturation : l'utilisation permanente des 12 notes de la gamme chromatique et des 24 tonalités possibles (12 tonalités majeures et 12 tonalités mineures) a contribué à faire disparaître toute hiérarchie entre les tons (cf. chromatisme de Wagner). Cette délitescence de la hiérarchie est à l'origine d'une perte progressive de toute conscience de tonalité chez l'auditeur. Cette crise est également celle de la mélodie, dont la perception est fondée sur la hiérarchie tonale. Après tout, la spécialisation sur l'intervalle n'était qu'une « *des facettes parmi bien d'autres de la prosodie du langage* » (Menezes, 1993 : 122). Cette crise était donc irréversible et devait déboucher sur une remise en cause complète des lois tonales. L'horizon des compositeurs se trouvait en effet complètement bouché s'ils persistaient dans la voie sans issue de la tonalité (qui avait déjà connu son apogée, donc n'offrait plus d'exploration possible). Parmi les alternatives possibles, on assiste à la création de nouveaux modes non fondés sur la tonalité (gammes par tons entiers, Debussy) et à la généralisation du chromatisme, qui va déboucher sur un bouleversement du langage musical, l'atonalité. Certains compositeurs veulent renouveler le matériau sonore et exploiter toutes les ressources que permet la voix humaine. Avec Debussy, on assiste à un retour à la musicalité intrinsèque de la langue française, c'est-à-dire à l'intonation. Mais, c'est Schoenberg qui va véritablement bouleverser les règles en inventant le Sprechgesang, nouveau type de déclamation à la frontière entre le chanté et le parlé.

2.2.1.2. Debussy et la musicalité de la langue française

Debussy veut libérer la mélodie de sa soumission à l'harmonie. Pour y parvenir, il va mettre au point une gamme par tons entiers où l'octave est divisée en 6 tons égaux. Dans une telle gamme, il n'y a plus de hiérarchie possible. Par ailleurs, il n'utilise plus les accords en fonction de la progression harmonique de l'œuvre, mais bel et bien pour leur beauté intrinsèque. Du coup, la mélodie ne peut plus être traitée en fonction des schémas traditionnels de tension et de détente. Debussy revient donc à la musique « naturelle » de la langue. Ainsi, dans son opéra « *Pelléas et Mélisande* » (1902), l'écriture musicale respecte le phrasé et l'accentuation de la langue du texte, le français. Le style vocal est très proche du récitatif et la mélodie connaît peu d'écarts d'intervalles. Dans l'extrait intitulé « *Devant le Château* » (Acte I, scène 3), Debussy met en scène la première rencontre entre Pelléas et Mélisande (**extrait sonore n°14**). Les personnages regardent le port et les bateaux s'éloigner, dont celui qui a transporté Mélisande jusqu'au château. La déclamation des chanteurs et l'orchestration suivent le rythme naturel du français : l'allongement final (figure n°42, syllabes accentuées indiquées en gras) et la rythmicité syllabique du français sont respectés. « *La fusion du son et du verbe est telle que texte et musique semblent n'avoir qu'un seul et même rythme* ». Quant à l'orchestration, elle suit même à certains endroits les inflexions de l'intonation : par exemple, pendant l'intervention de Mélisande, « *Je ne sais si nous pourrons le voir, il y a encore une brume sur la mer* ». Debussy fait chanter ses personnages « *comme des personnes naturelles et non pas dans une langue arbitraire, faite de traditions surannées* ». Les voyelles ne sont plus privilégiées comme elles l'étaient dans la musique tonale. La ligne mélodique s'y attardait en faisant fleurir des ornements. Mais, à partir du moment où on revient à la musique « naturelle » de la langue, on retrouve l'équilibre des sons propre à cette langue, notamment les voyelles et les consonnes. L'auditeur est ainsi placé devant la musicalité de sa propre langue (cf. Menezes, 1993 : 89). Debussy conserve cependant quelques petites différences : le e caduc ne chute pas en finale (par exemple « *haut-es* » et « *navi-re* ») comme il devrait le faire dans la langue parlée. Ce maintien engendre un dédoublement de la syllabe et permet au compositeur d'introduire une note supplémentaire, qui « musicalise » un peu plus le phrasé.

M : Quelque-cho-se sort du **port**.
 P : Il faut que ce soit un grand **navi**-re, les lumiè-res sont très **hau**-tes. Nous le verront tout-à-**l'heure**, quand il entrera dans la voile de **clarté**.
 M : Je ne sais si nous pourrons le **voir**, il y a encore une **bru**-me sur la **me**-re.
 P : On dirait que la **bru**-me s'élè-ve **lentement**.
 M : Et j'aperçois là **bas** une petite **lumiè**-re que je n'avais pas **vue**.
 P : C'est un **pha**-re. Il y en a **d'au**-tres que nous ne voyons pas **encore**.
 M : Le **navire** est dans la **lumiè**-re ; il est déjà bien **loin**.
 P : Il s'éloigne à toutes **voiles**.

Figure n°42: accentuation de l'extrait « Devant le Château », correspondant à l'extrait sonore n°14, « Pelléas et Mélisande », de Debussy.

2.2.2.2. Le « Sprechgesang »

Nous l'avons vu, l'évolution naturelle du système tonal le pousse inéluctablement vers la saturation et l'une des possibilités offertes aux compositeurs consiste à généraliser l'utilisation du chromatisme dans l'écriture. C'est ainsi qu'Arnold Schoenberg va créer la **musique dodécaphonique**. Afin d'éviter la formation de centres tonals, il met au point une technique sérielle où chacun des 12 demi-tons ne réapparaîtra pas avant que les 11 autres ne soient apparus (variation continue). Ce principe est très contraignant pour la composition car seule la répétition immédiate d'une même note est permise. Mais, grâce à cette technique sérielle, toutes les tonalités peuvent coexister de manière simultanée. Comme Debussy, Schoenberg doit repenser la mélodie, car celle-ci ne peut plus exister sous sa forme traditionnelle dans un système atonal. Comme lui, il va revenir à la musicalité de la langue, mais il ne va pas l'utiliser de la même façon. Dans son « *Pierrot Lunaire* » (1912), il met au point un mode de déclamation qui se situe à la frontière du chanté et du parlé : le « Sprechgesang ». A cause de son étrange statut, on pourrait le traduire indifféremment par « parole chantée » ou par « chant parlé ». En effet, il est difficile de déterminer s'il tient plus du chanté que du parlé et inversement. Les limites habituelles entre langue et musique sont transgressées d'une manière beaucoup plus forte que dans les formes de récitatif précédentes. Le procédé de notation adopté par Schoenberg traduit cette fluctuation du Sprechgesang entre la langue et la musique : le rythme est donné, mais en revanche, les hauteurs ne sont pas indiquées de manière précises (une croix remplace le cercle de la note, pour suggérer que la hauteur n'est pas fixée, qu'elle est plus haute ou plus basse que celle qui est indiquée par la croix). Par ailleurs, avant Schoenberg, la musique vocale se limitait somme toute à deux types d'émission opposés : le parlé et le chanté, alors qu'il existe toute une palette d'intermédiaires entre ces deux extrêmes.

Schoenberg, lui, ne se limite pas à ces deux modes de production vocale : en plus du parlé et du chanté, il utilise également les rires, les soupirs, les chuchotements, les gémissements, les plaintes... Le Sprechgesang constitue une réelle tentative d'exploration de toutes les possibilités d'expression qu'offre la voix. Non seulement Schoenberg les explore, mais surtout il les mélange. Ainsi, selon Gagnard (1987 : 14),

« il introduit volontiers un son de hauteur précise chanté au milieu d'une phrase en parlando. Toute cette technique de Sprechgesang est basée sur les élans et les retombées, mais ce n'est pas aussi simple, car il mélange constamment les genres d'émission, par exemple le rire (à demi) et le chanté ».

A titre illustratif, nous avons choisi un extrait non pas du « *Pierrot Lunaire* », mais de « *Ode to Napoleon Buonaparte opus 41* », (1942) (**extrait sonore n°15**), car le Sprechgesang y est très affirmé. Le texte en anglais provient d'un texte de Lord Byron (voir texte et traduction en français, figure n°43). La mélodie et le rythme principal sont assurés par la voix du baryton : c'est lui qui conduit l'interprétation. Il n'y a plus de pulsation rythmique. Dans ce passage, la musique n'a plus qu'une fonction d'accompagnement : les cordes suivent, ponctuent et illustrent la « narration ». Le début de l'extrait oscille entre le parlé (« *than littleness* »), le scandé (« *ambitions* ») et le légèrement chanté (« *mortals guess* »). On trouve des phrases entièrement parlées, comme « *it will **teach to after-warriors more than high philosophy can preach, and vainly preach'd before*** » avec un respect scrupuleux de l'accentuation et de l'intonation anglaise (accent sur la deuxième syllabe de « *philosophy* » par exemple). Un peu plus loin, le rythme est tellement scandé sur la phrase « *Those **pagod things of sabre way*** », qu'une pulsation est réintroduite. Cet extrait est suivi quelques mots plus tard d'une note chantée avec une hauteur fixée sur « *Brass* », puis d'un passage parlé avec « *and feet of clay* ». La suite consiste en une pure déclamation théâtrale « *The triumph and the vanity, the rapture of the strike* », qui englobe le début de la phrase suivante « *the earthquake voice of* », pour basculer brutalement dans du chanté sur le dernier mot « *Victory* », et pas n'importe quel chanté. En effet, le piano a commencé quelques instants plus tôt une citation musicale évoquant le début du premier mouvement de la *Cinquième Symphonie* de Beethoven et la voix le rejoint sur « *Victory* ». On retrouve l'espace d'un mot le statut traditionnel de la voix chantée. Sans arrêt, Schoenberg alterne les modes vocaux, de manière à ne jamais laisser la monotonie s'installer, l'intrusion de passages chantés dans le parlé en est un exemple. L'oreille de l'auditeur doit toujours rester en alerte, sans jamais pouvoir anticiper ce qui va suivre.

Ce mélange continu contribue encore davantage à abolir les limites entre le parlé et le chanté.

Nor till thy fall could mortals guess Ambitions less than littleness ! Thanks for that lesson – it will teach To after-warriors more Than high philosophy can preach, And vainly preach'd before. That spell upon the minds of men Breaks never to unite again, That led them to adore Those pagod things of sabre way, With fronts of brass, and feet of clay. The triumph, and the vanity, The rapture of the strike – The earthquake voice of Victory, To thee the breath of life ; The sword, the sceptre, and that sway Which man seem'd made but to obey Wherewith renown was rife – All quell'd ! –	C'est depuis ta chute que les mortels connaissent Que l'Ambition est moindre que la petitesse ! Pour la leçon, merci – car par elles apprendront Les guerriers de l'avenir Plus que des hauts prêches de la Philosophie, Qu'avant elle prêchait en vain. Ce charme qui pesait sur les esprits des hommes Est brisé pour jamais pouvoir se restaurer, Qui les menait à adorer Ces êtres fabuleux de l'empire du sabre, Dont les fronts sont d'airain, dont les pieds sont d'argile. Le triomphe et la vanité, dans le combat Le ravissement, et la voix De la Victoire tel un tremblement de terre, Qui te furent souffle de vie ; Et le glaive, et le sceptre, et ce pouvoir par quoi L'homme ne semble fait rien que pour obéir, Dont le renom régnait partout, Tout cela est dompté !
--	---

Figure n°43 : texte en anglais et traduction en français de l'extrait de « Ode to Napoleon Buonaparte, opus 41 » correspondant à l'extrait sonore n°15.

2.2.3. Musicalisation de la parole dans la création contemporaine

« Une attitude plus ambitieuse consiste à proposer entre toutes, la musique comme une activité « globalisante », comme une interdiscipline proprement dite, une activité qui, recoupant de multiples disciplines spécifiques, vérifie par synthèse leurs apports partiels, tant sur le plan des faits que sur celui des idées, et se présente au même titre qu'elles comme une activité de découverte, qui vise autant, sinon plus, à fonder une connaissance qu'à créer des œuvres », Pierre Schaeffer (1966 : 31).

2.2.3.1. La langue traitée comme matériau

Selon l'Encyclopedia Universalis, la notion de matériau recouvre en musique

« tout ce qui sert de matière première à l'écriture, c'est-à-dire l'harmonie, l'intervalle, le timbre, le rythme, la forme » (Fréron, 2002).

Or, toute cette matière première est élaborée à partir de la combinaison des paramètres acoustiques que sont la hauteur, la durée, l'amplitude et le timbre. Ces paramètres se trouvent agencés selon une syntaxe spécifique, qui était celle de la tonalité jusqu'au XXe siècle. La remise en cause de cette syntaxe par des compositeurs comme Schoenberg et Berg entre autres a permis la naissance de l'atonalité. Au milieu du XXe siècle, une autre révolution va venir complètement modifier le rapport du compositeur au matériau musical, la naissance de l'électro-acoustique. Grâce à elle, de nouveaux instruments voient le jour comme les ondes Martenot par exemple en 1928. Certains instruments traditionnels s'électrifient et sont amplifiés (guitare, violon, etc.). Mais, ce sont surtout les techniques d'enregistrement et d'analyse du son qui connaissent des progrès considérables, avec, pour conséquence, une modification directe du rapport entre le compositeur et le matériau. Celui-ci peut enfin enregistrer, manipuler les sons, les découper et les isoler, les coller et les assembler, les modifier, les créer ex nihilo... Il peut ainsi se transformer en ingénieur du son, ce qui implique un rapport complètement différent à la musique. Alors qu'il avait besoin d'un orchestre pour la mise en musique de ses compositions, il peut dorénavant créer ses œuvres au sein même du studio. Il n'a plus forcément besoin d'un orchestre, mais peut réutiliser des enregistrements ou incorporer un instrument électronique à un orchestre traditionnel (utilisation de potentiomètres par exemple). Ainsi, Pierre Schaeffer, qui est également ingénieur, utilise des sons concrets, c'est-à-dire des sons enregistrés qu'il assemble dans ses compositions ; quand à Edgard Varèse, il sera le premier à coupler des sons « organisés » à un orchestre dans une œuvre qui scandalisera le public, « *Déserts* » (1954).

Grâce à ces techniques, la parole jusque là éphémère, se transforme en objet fixé, donc reproductible, que l'on peut transformer à volonté par le découpage, le collage et la synthèse. Autrement dit, on peut lui faire subir toutes les distorsions que l'on veut. Cette possibilité est en soi une petite révolution, car elle entraîne un changement de statut de la langue. En passant de la catégorie des « intouchables » à la catégorie des « transformables », la langue devient accessible, elle devient un simple matériau. Ce changement de statut entraîne de nouvelles réflexions sur la langue et la musique et spécialement sur les rapports qu'elles entretiennent. Toutes deux peuvent en effet servir de matériau au compositeur qui peut les manipuler comme bon lui semble. En partageant le même statut, elles deviennent solidaires jusqu'au point où dans certaines œuvres comme celles de Pierre Boulez et Luciano Berio, l'opposition musique/langage se voit supprimée.

2.2.3.2. Neutralisation de l'opposition langage/musique chez Berio

Debussy avait largement amorcé le retour à l'intonation dans Pelléas et Mélisande, mais pour Menezes (1993 : 89), c'est Luciano Berio qui

« constitue peut être l'exemple le plus singulier de toute sa génération d'une prise de conscience radicale de ce retour à l'origine de la musique en tant qu'intonation de la parole ».

La « *Sequenza III* » pour voix de femme a été créée par Berio en 1966 au Studio di Fonologia de la RAI à Milan. Le texte en anglais de Markus Kutter a perdu toute intelligibilité lors de sa mise en musique. Il est très difficile de l'identifier à moins de l'avoir sous les yeux : « *Give me a few words for a woman to sing a truth allowing us to build a house without worrying before night comes* ». Berio procède en effet à une instrumentalisation de la voix, en découpant les mots en syllabes, en phonèmes. Ainsi, « *la parole est secouée jusqu'à ses racines phonétiques* » (Berio, cité par Gagnard, 1987 : 49). L'**extrait sonore n°16** est une caricature de la parole grâce à la parole elle-même. La voix prononce un discours en charabia, ponctué par des notes aiguës et intenses, puis s'élève dans un simulacre de rire. L'extrait se termine par une suite de trois notes tenues, dont la première et la troisième sont émises bouche ouverte et la deuxième bouche fermée. Le texte est découpé, pour être totalement recomposé au sein d'une œuvre musicale originale.

« Pour contrôler un ensemble aussi vaste de comportements vocaux, il a fallu morceler le texte et le détruire en apparence, afin de récupérer les fragments sur divers plans expressifs et recomposer l'ensemble en unités non plus discursives mais musicales » (Berio, 1998, livret d'accompagnement du CD).

Du coup l'opposition musique/langage n'existe plus, car la parole est complètement musicalisée. Cette absence d'opposition se reflète dans la partition où la notation musicale cohabite avec la transcription phonétique (API) et la transcription orthographique (figure n°44).

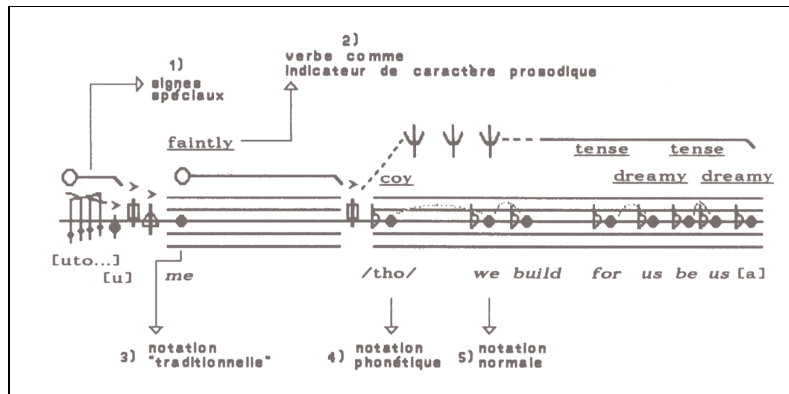


Figure n°44 : extrait d'une transcription « Sequenza III » (page 1) (emprunté à Menezes, 1993 : 124).

Par ailleurs, Berio a recours au pouvoir expressif des mots pour indiquer les caractères prosodiques que le chanteur doit appliquer au texte (figure n°45) : « anxieux », « tendu », « noble » ou encore « extatique ». Bien évidemment, la notation musicale ne peut transmettre de tels états. Le problème, c'est que la notation est surchargée et très difficile à déchiffrer par les interprètes. Ce manque de lisibilité pose les limites de la notation musicale classique, aussi bien que celles de la transcription orthographique et phonétique. Alors que les phonéticiens ont tenté de trouver des systèmes de transcription capables de restituer la prosodie (partie 2.1.1.2.), il est frappant de constater que la musique contemporaine se heurte au même problème, dans la mesure où elle revient à l'intonation et au rythme langagier. Cette problématique commune résulte de l'effacement des limites entre langue et musique.

Exemple 1a: caractères prosodiques dans "Sequenza III"

anxious	joyful
apprehensive	languorous
bewildered	muttering
calm	nervous
coy	nervous L.
distant	noble
distant and dreamy	open L.
dreamy	relieved
dreamy and tense	serene
echoing	subsiding
ecstatic	tender
extremely intense	tense
extremely tense	tense L.
fading	tense muttering
faintly	urgent
frantic	very excited and fran
frantic L.	very tense
gasping	walking on stage
giddy	whimpering
impassive	whining
increasingly desperate	wistful
intense	witty

Figure n°45 : caractères prosodiques dans la « Sequenza III » de Berio (1966) (emprunté à Menezes, 1993 : 123).

2.2.3.3. Les « musiciens phonéticiens »

Les techniques de traitement du son n'ont pas bénéficié uniquement à la musique, mais également aux sciences phonétiques et il existe dorénavant une sphère commune entre le musicien et le phonéticien, par l'intermédiaire de l'analyse du signal. Par conséquent, il n'est guère étonnant que beaucoup de compositeurs se soient intéressés à la linguistique, par exemple Berio à la phonologie. Dans la création contemporaine, on trouve même des musiciens tels que Claudy Malherbe et Luca Francesconi qui fondent leurs compositions sur une analyse phonétique : ils musicalisent la parole au sein d'une nouvelle discipline que l'on pourrait s'amuser à surnommer la phonémusique. Mais, si beaucoup de musiciens se passionnent pour la langue, la réciproque n'est pas forcément vraie et les travaux réalisés par des phonéticiens (phonéticiens-musiciens) sur les rapports entre langue et musique restent encore peu nombreux (à l'exception notable des travaux de Fonagy sur l'expression des émotions dans la langue et dans la musique).

L'IRCAM¹⁵, fondé par Pierre Boulez en 1975, est un lieu de recherche multidisciplinaire sur les techniques de composition contemporaines. Le programme « *Les noces de la voix et de l'électronique* »¹⁶ regroupe des compositeurs qui travaillent à l'extension des techniques vocales. Parmi eux, Claudy Malherbe compose l'harmonie et la mélodie de « *Locus* » (23 juin 1997, Espace de projection de l'IRCAM) pour quatuor vocal et dispositif électronique, à partir de l'analyse acoustique de phrases en anglais. Il a préféré l'anglais au français à cause de sa musicalité intrinsèque (plus d'accentuation et variations mélodiques plus étendues). Les phrases qui composent le texte sont tirées de manuels de phonétique et d'acoustique,

*« l'ensemble fournissant à la fois un matériau vocal ainsi qu'un commentaire qui donne, au fur et à mesure du déroulement, la description des éléments et des situations en jeu (l'appareil vocal : larynx, gorge, langue, lèvres... ; respirer, parler, manger »*¹⁷.

L'écriture musicale a été élaborée directement à partir de l'analyse acoustique et phonétique de ces phrases. Le compositeur a procédé de la manière suivante : il a commencé par segmenter les phrases en syllabes, puis a réalisé un lissage en moyennant les différents partiels à l'intérieur de chaque syllabe. Il obtient ainsi des bandes de fréquence amplifiées dont l'évolution est statique et les phases de transition effacées.

¹⁵ IRCAM : Institut de Recherche et de Coordination Acoustique/Musique, Centre Georges Pompidou, Paris.

¹⁶ Programme exposé sur le web : <http://sol.ircam.fr/instruments/musicweb.fr>

¹⁷ Commentaires et présentation de l'œuvre de Claudy Malherbe disponibles sur le site web de l'IRCAM.

Toutes ces bandes sont converties en valeurs musicales de manière à former un accord (ou agrégat) correspondant à chaque syllabe. Ces accords sont ensuite reportés sur une double portée musicale (clé de fa/clé de sol). L'harmonisation de « *Locus* » est donc établie à partir de la répartition spectrale de la parole naturelle.

Dans « *Etymo* »¹⁸ (25 novembre 1994, Espace de projection de l'IRCAM, Paris) pour soprano, ensemble et dispositif électronique, Luca Francesconi utilise des extraits de poèmes de Baudelaire (« *Le Voyage* », « *L'Albatros* » et « *Carnets Intimes* », 1862). Pour lui, le mot est musique ; il est donc allé recueillir cette musique en procédant lui aussi comme un phonéticien : il a demandé à la soprano de lire ces textes, puis il a réalisé une analyse spectrale des voyelles et des consonnes, ainsi qu'un relevé de la durée des pauses et des différents segments. « *Etymo* » a été composé à partir de ces mesures. Le texte de l'extrait présenté (**extrait sonore n°17**) est fondé sur le retour des sons [μ] et [ε] : « *Même dans nos sommeils, la curiosité nous tourmente* » (partition, **figure n°46**). Les accords sont composés à partir de la répartition spectrale des voyelles.

Figure n°46 : extrait de « *Etymo* » de Luca Francesconi (1994).

2.2.3.4. Typologie des effets : « langue dans la musique »

La musique contemporaine n'est pas un champ clos, dont les expérimentations et les avancées resteraient cantonnées à sa propre sphère d'influence, à ses propres compositeurs et à son propre public.

¹⁸ Commentaires et présentation de l'œuvre de Lucas Francesconi, disponibles sur le site web de l'IRCAM.

Elle exerce son ascendant sur d'autres courants musicaux, plus « populaires ». Ce fut le cas durant les années 60-70 avec la musique pop, qui fut également un terrain d'expérimentation très fertile pendant cette période. Ainsi, un musicien comme Franck Zappa était fasciné par les compositions d'Edgard Varèse et il réalisa beaucoup d'expérimentations, notamment avec son groupe, les *Mothers of Invention*. L'électroacoustique fut largement utilisée et donna naissance à l'instrument phare de la pop : la guitare électrique. Aujourd'hui, la musique électronique fournit également un champ d'investigation fructueux. Il est donc très important de ne pas se cantonner à un seul type de musique, ni bien sûr à une seule culture musicale. Cependant, plutôt que de réaliser une étude qui fasse un inventaire complet sur le sujet et qui serait trop peu en rapport avec notre sujet, il semble plus pertinent d'établir une typologie des différents procédés qui permettent d'incorporer la musique de la langue dans la musique. Cette typologie ne se prétend pas exhaustive, mais jette les bases pour la réalisation d'un catalogue plus complet. Elle est constituée de trois grandes classes : la première englobe toutes les utilisations « naturelles » de la langue dans une composition musicale, imitation ou incorporation sans modifications du donné langagier ; la deuxième comprend toutes les compositions musicales élaborées à partir d'une combinaison spécifique d'éléments prosodiques. La troisième concerne les compositions élaborées à partir d'une modification du signal avec découpage, collage et/ou synthèse (modification par traitement du signal sonore).

Utilisations naturelles de la langue dans une composition musicale :

a) Imitation de la prosodie d'une langue naturelle par un instrument

Les musiciens imitent l'intonation avec leur instrument. Nous en avons déjà donné un exemple avec la guitare parlante de Jimi Hendrix (**extrait sonore n°1**). Cet effet est utilisé par Robert Wyatt lorsqu'il se met à dialoguer avec son saxophone dans « *Alife* » (**extrait sonore n°18**). On le retrouve également dans les improvisations de jazz (Duke Ellington) et de blues (Luther Allison). De la même façon, dans le spectacle « *Con moto* » de La Framboise Frivole, le musicien belge Peter Hens dialogue avec son violoncelle. En musique classique instrumentale, on trouve un passage qui a été traité comme du chant parlé : le violoncelle et la contrebasse entament un récitatif comme le ferait une voix humaine au début de l'Ode à la Joie de Schiller, dans la Neuvième Symphonie de Beethoven dont nous avons déjà parlé plus haut (**extrait sonore n°13**).

Enfin, nous ne pouvions clore ce paragraphe sans parler du phonéticien Kenneth Pike et de sa flûte. La coulisse de cet instrument lui permettait en effet d'imiter les modulations de l'intonation.

b) Incorporation de clichés mélodiques

Bobby McFerrin incorpore le cliché mélodique « *na-na-na-na-nè-re* » dans son improvisation vocale (**extrait sonore n°4**). Les clichés sont souvent utilisés dans les improvisations, en raison de leur musicalité, mais surtout parce qu'ils sont immédiatement reconnus par les membres d'une même communauté linguistique. Ce faisant, le musicien adresse une sorte de clin d'œil musical à l'auditoire.

c) Utilisation de la musicalité naturelle de la langue sans modifications

La parole possède une musicalité intrinsèque, une mélodie et un rythme spécifique que les musiciens peuvent réutiliser telle quelle, sans la modifier. Au début de « *Kids can dance* », le groupe Hexstatic utilise l'intonation d'un énoncé en anglais comme ligne mélodique : « *I know you like music, Here's again, It is yours, I canno't accepted, I want you to have it, or before you go, please, play that music for me. Yes master* » (**extrait sonore n°19**). Le rythme d'une batterie vient se superposer au rythme et à la mélodie « naturelle » de cette phrase, sans les modifier.

d) Composition fondée sur la musicalité naturelle de la langue

La composition peut s'inspirer de la musicalité naturelle de la langue. Ainsi, nous avons vu que le chant grégorien était fondé sur la musicalité du latin (**extrait sonore n°5**). De même, le récitatif donne la priorité au texte, donc à la musicalité de la langue de composition (**extraits sonores n°7, 9 et 11**). Dans *Pelléas et Mélisande*, Debussy a élaboré son écriture à partir de la musicalité de la langue française (**extrait sonore n°14**).

Incorporation de la prosodie à l'écriture musicale et élaboration d'un matériel original

e) Mode de déclamation à la frontière du parlé et du chanté

Le Sprechgesang est une technique de déclamation qui exploite l'ambiguïté entre la langue et la musique, qui mélange le parlé et le chanté d'une manière telle qu'il devient difficile de déterminer ce qui appartient à la langue et ce qui relève de la musique (**extrait sonore n°15**).

f) Polyphonie d'un petit groupe de voix, structurées autour du retour de certains événements phonétiques (Crystal, 1987 : 173)

Dans le morceau « *It can't happen here* », Franck Zappa réalise avec les autres membres de son groupe une polyphonie structurée sur le retour de faits phonétiques et prosodiques (**extrait sonore n°20**). La phrase introductive de l'extrait est chantée par Zappa lui-même : « *But they will freak out in Wahinsgton D.C.* ». Une polyphonie rythmique lui succède qui se structure de la manière suivante : l'accompagnement est composé de phrases scandées par les différents membres du groupe. Ces phrases sont répétées et alternées sans interruption de manière à former une rythmique stable, dont on peut extraire la pulsation : : « *DC, DC, DC* » + « *they can't happen here* » + « *ba ba ba* »... Un peu plus loin vient se greffer la « mélodie principale », déclamée par Zappa : « *everybody say that can't happen here ; no freak for us, can't happen here ; everybody's clean and can't happen here, no, no, it won't happen here, I tell you ...* ». Dans un autre passage (0 mn 48), l'énoncé suivant est scandé par un chanteur : « *And they thought it couldn't happen here ; and knew it couldn't happen here. They were so sure it couldn't happen here, but...* ». Les autres voix l'accompagnent en interprétant des parodies de l'intonation de cette phrase sur la suite déphonématisée « *mamama...* ».

g) Instrumentalisation de la voix

Dans la *Sequenza III* (**Extrait sonore n°16**), Luciano Berio instrumentalise la voix en découpant les mots en syllabes et en phonèmes, qu'il recompose ensuite à sa guise pour créer une musique originale. C'est une sorte de collage au niveau de la composition, sans avoir recours à une manipulation du signal sonore.

Traitement du signal sonore (montage- collage – analyse)**h) Reconstitution d'une mélodie originale à partir du collage de plusieurs fragments langagiers**

Dans « *Ninja tune* », le groupe Hexstatic¹⁹ crée une mélodie à partir du collage de plusieurs mots : « *Media* », « *one* », « *solo* » et « *deadly* », prélevés au sein de différents énoncés et prononcés par différentes voix (**extrait sonore n°21**). L'assemblage de ces mots au sein d'un nouvel énoncé « *Media, one, one, one solo, media, one, one, one, deadly* » crée une nouvelle mélodie. La rythmique est générée par l'agencement spécifique de ces mots (première phrase de l'extrait) auquel vient s'ajouter à la phrase suivante un rythme de batterie.

i) Effet cocktail : polyphonie de grands groupes (Crystal, 1987 : 173)

Dans « *Ninja tune* », le groupe Hexstatic réalise une polyphonie en exploitant l'effet cocktail. Différentes voix d'hommes et de femmes, s'exprimant dans différentes langues (japonais, anglais, espagnol, allemand...), ont été prélevées dans des journaux radiophoniques. Ces voix ont ensuite été réagencées de manière à créer une polyphonie (**extrait sonore n°22**). L'extrait commence par une voix d'homme qui s'exprime en japonais. Une autre voix d'homme en anglais vient s'y superposer, suivie quelques instants plus tard d'une voix de femme en espagnol. A mesure que le nombre de voix se rajoutent, les paroles deviennent inintelligibles et l'oreille ne perçoit plus qu'un brouhaha, une bouillie sonore, de laquelle émerge progressivement à la fin de l'extrait une rythmique introduite par une basse, à laquelle vient s'ajouter un « montage » rythmique réalisé à partir du collage de différents fragments prélevés dans les phrases (voir suite, h, **extrait sonore n°21**). Grâce au collage, on peut imaginer un autre type de polyphonie, l'autopolyphonie où le montage est réalisé à partir de différents enregistrements d'une seule et même voix (Crystal, 1987 : 173).

¹⁹ Site web du groupe Hexstatic : <http://hexstatic.co.uk/>

j) Écriture musicale réalisée à partir d'une analyse acoustique du donné langagier

Des compositeurs comme Luca Francesconi (**extrait sonore n°17**) et Claudy Malherbe fondent leur écriture musicale sur une analyse acoustique de la parole. Les accords musicaux sont par exemple déduits de la répartition spectrale des voyelles chez Francesconi.

3. Similitudes de formes étendues à des similitudes structurales

3.1. Universaux

Langue et musique ont de fortes similitudes de formes, à tel point que dans certaines situations, nous avons vu qu'il était difficile de les distinguer l'une de l'autre. Il n'est donc pas étonnant que ces similitudes soient exploitées par les spécialistes de la matière sonore que sont les phonéticiens et les musiciens. Nous ne serions toutefois pas complets si nous n'abordions pas le travail des chercheurs d'autres disciplines. Il est en effet très séduisant d'étendre ces similitudes de forme à des similitudes de structure et de chercher à savoir si langue et musique ne partagent pas des « structures profondes » similaires, voire identiques. Ainsi, beaucoup de théoriciens en linguistique ou en musicologie ont tenté de définir des universaux communs à la langue et à la musique. Ils est vrai qu'elles ont toutes les deux un statut particulier dans l'histoire de l'humanité : elles se sont développées dans toutes les sociétés humaines et cela, quelle que soit l'avancée technologique atteinte par chacune de ces sociétés. Tous les êtres humains sont capables de parler et de faire de la musique, sans apprentissage préalable. L'acquisition de la langue, comme celui de la musique se fait de manière « naturelle », par simple confrontation avec le milieu (cf. chapitre II). Et la variation entre les différents milieux est à l'origine aussi bien de la diversité linguistique, que de la diversité musicale. La langue et la musique constituent donc des universaux au sein des pratiques humaines. Une connaissance approfondie de ce qui rapproche langue et musique permet de mieux cerner leurs différences, nous l'avons vu dans la partie 1.2. pour le rythme et la mélodie. Et peut-être de relever le défi lancé par Mc Adams, « *comprendre en quoi la musique est complètement différente de tout autre chose* », défi qu'on peut lancer pour la langue également. Quelles qu'en soient les conclusions, une telle approche comparative ne peut que se révéler riche d'enseignement.

Cette approche se montrera d'autant plus riche que le chercheur aura pris soin d'éviter certains pièges, dont le plus grave serait de caricaturer l'un des systèmes à travers le filtre de l'autre et de lui appliquer des concepts qui ne lui conviennent pas.

3.2. Similitudes structurales

Les messages verbaux et musicaux ont en commun d'être construits à partir de la **combinaison d'un certain nombre d'unités sans signification**, les phonèmes pour la langue (26 pour le français) et les notes pour la musique (12 pour la gamme chromatique). Dans les deux systèmes, ce **répertoire d'unité** est **limité**. Mais ces unités de bas niveau, sans signification propre peuvent se combiner entre elles pour produire des unités plus complexes à un niveau plus élevé, qui elles-mêmes vont se combiner, pour atteindre un niveau encore plus élevé. Ainsi, selon Lerdahl et Jackendoff (1983), un énoncé se divise en phrases, ces phrases se divisent en propositions, les propositions en mots, les mots en syllabes et les syllabes en phonèmes ; une pièce musicale se divise en sections, ces sections se divisent en phrases musicales, les phrases en figures mélodiques et les figures rythmiques en notes. Ces unités de différents niveaux répondent donc à une **organisation hiérarchique**. Cette structuration donne la possibilité à la langue et à la musique de créer une infinité d'énoncés porteurs de signification, alors qu'elles partent toutes les deux d'un répertoire très restreint d'unités de bas niveaux. Elles sont donc de **nature générative**. Cette structuration hiérarchique de la langue et de la musique n'est pas seulement une abstraction, mais elle prend une forme très concrète grâce au marquage acoustique de certaines unités (voir chapitre II, partie 2.3.1.). Ainsi, la prosodie contient de nombreux indices sur les frontières linguistiques importantes. Par exemple, la frontière entre les propositions est marquée par une augmentation ou une diminution de la Fo ou par un allongement de la syllabe précédant la frontière de la proposition (Klatt et Cooper, 1975). De la même manière, les frontières musicales sont indiquées par une chute de la hauteur et un allongement de la durée des notes (Juszyk et Krumhansl, 1993). Ce marquage acoustique des constituants va se révéler d'une importance fondamentale dans l'acquisition de la langue et du système musical maternel par l'enfant au cours de la première année de sa vie (voir chapitre II).

3.3. Peut-on parler de grammaire et de langage à propos de musique ?

La linguistique générative définit la grammaire comme un ensemble fini de règles qui rendent compte d'un nombre illimité d'énoncés différents. S'il est possible d'établir la grammaire d'une langue grâce à la description de ces éléments constitutifs et des règles qui permettent de les combiner, il devrait donc être possible d'établir la grammaire d'une musique spécifique, d'autant que nous venons de voir que la musique, comme la langue obéit à une organisation hiérarchique d'unités. Pour un auteur comme Chomsky, la grammaticalité de la musique ne fait aucun doute. À vrai dire, cette idée n'est pas neuve et il existe déjà des grammaires de la musique modale, de la musique tonale ou de la musique sérielle par exemple. Les musiciens doivent apprendre et maîtriser ces systèmes de règles s'ils veulent pouvoir composer dans ces « langages musicaux ». Mais selon l'approche générative, la grammaire ne consiste pas seulement en une description des unités et des règles couchée sur le papier, elle est surtout et avant tout un système intériorisé par l'auditeur, qui lui permet d'interpréter les messages qu'il reçoit et d'en générer par lui-même. Cette grammaire intériorisée lui donne la capacité de juger si une séquence est correcte ou non, autrement dit, si elle est grammaticale ou non. Cette « compétence » a été décrite pour la langue par la grammaire générative et transformationnelle de Chomsky. Mais, existe-t-il une compétence similaire pour la musique ? Pour le savoir, Sloboda (1976) a demandé à des pianistes de déchiffrer des pièces musicales inconnues. Ces pièces respectent scrupuleusement les règles du langage musical dans lequel elles ont été composées (musique tonale), mais comportent des erreurs de notation. Les règles sont tellement intériorisées par les pianistes, qu'ils se montrent totalement incapables de détecter ces erreurs, car ils les corrigent inconsciemment au moment de la lecture. La grammaire de la musique tonale semble donc bien être intériorisée chez l'auditeur, qui développe ainsi, comme pour sa langue maternelle, une « compétence » à juger si les énoncés musicaux sont corrects ou non, s'ils sont grammaticaux ou agrammaticaux.

Si la musique possède une grammaire, alors, peut-on parler de langage à propos de la musique ? En fait, tout dépend de la définition du langage qu'on adopte. Pour Cooke (1959), ce terme désigne simplement

« un système de communication d'un état émotionnel entre le compositeur et les auditeurs, lequel système incorpore une sorte de lexique informel d'unités musicales expressives » (Clarke, 1989 : 24).

Si l'auditeur maîtrise le langage de la musique, il pourra comprendre le message que le compositeur lui adresse par le biais du canal que représente son œuvre ; s'il échoue, il sera considéré comme peu musicien. Cette transmission de l'intention du compositeur à l'auditeur est rendue possible selon Cooke grâce aux correspondances très stables qui existent entre la configuration mélodique et le contenu affectif dans la musique occidentale tonale. Osmond-Smith (1989 : 145) critique vertement ce type de théorie où l'auditeur est censé comprendre le message que le compositeur veut lui faire passer :

« cette sorte de modèle communicatif s'applique assez bien au morse, mais je ne suis pas vraiment convaincu qu'il décrive de façon adéquate le processus impliqué dans un médium artistique ».

Selon Ruwet (1972 : 44), il ne faut tenir compte de ce que nous a appris la linguistique et ne pas s'en tenir à la conception « langage-signal ».

Pour Pierre Schaeffer (1966), si la musique est un langage, ce n'est pas au même titre que le Langage proprement dit car la structure musicale se trouve liée indissolublement aux qualités sensibles de son matériau :

« on peut mettre en doute un parallélisme étroit entre langue et musique, en raison de l'arbitraire qui reste attaché au choix du sens, de la relation libre du signifiant et du signifié, qui fait du mot un signe, alors que la note de musique a toujours paru s'imposer en dehors de tout arbitraire comme une donnée du monde physique, à la quelle nous serions sensibles » (Schaeffer, 1966 : 35).

D'après lui, dans toute l'histoire de la musique, c'est la « musique pure » qui se rapproche le plus de la langue, langue dans le sens que lui donne Saussure, c'est-à-dire dégagée de la parole. Il prend pour exemple de musique pure l'« *Art de la Fugue* », composée en 1749 par Jean-Sébastien Bach. Cette œuvre a été écrite sans prévoir aucune instrumentation et Chion souligne que

« la musicalité est entièrement dégagée de la sonorité et l'objet sonore y est entièrement utilisé comme signe ; il n'y a que des relations et la musique n'est plus faite que d'un jeu de valeurs, de traits distinctifs, de hauteurs et de durée » (Chion, 1983 : 83).

Seule la hauteur peut constituer selon Schaeffer la valeur de base d'une musique pure, grâce à sa haute capacité d'abstraction et notamment sa capacité d'être abstraite des objets qui la supportent :

« la musique traditionnelle tend à s'abstraire peu à peu de la sonorité pour se constituer en langue, c'est-à-dire en système de valeurs pures [...] et de la sorte, à ne retenir des objets sonores que tel ou tel trait distinctif, compromis entre une acoulogie naturelle et sociale. Tel est le point où musique et langage se rapprochent le plus et en donnent des preuves évidentes » (Schaeffer, 1966 : 309).

Seulement, cette proximité ne peut être atteinte que si le langage musical respecte le cadre naturel des rapports simples de quinte et d'octave, donc qu'il se fonde sur des critères de perception musicale. La musique pure n'échappe pas à la loi qui veut que toute musique soit faite *« pour être entendue »* et c'est justement cette caractéristique qui constitue la différence entre langue et musique : alors que la signification du langage n'a qu'un lien arbitraire avec son support acoustique, les valeurs musicales, mêmes abstraites, sont reliées aux propriétés de l'objet. La musique est en effet bâtie sur un système où le signe n'est pas arbitraire et où il fonde ses valeurs sur les propriétés effectives de l'objet perçu, c'est-à-dire les rapports simples de hauteurs.

Seulement, lorsque Schaeffer parle d'arbitrarité du signe, il se réfère à la relation entre le signifiant et le signifié exposée par Saussure et cette définition ne tient justement pas compte de la prosodie, elle s'applique seulement au mot. Elle a d'ailleurs contribué à marginaliser la prosodie au sein des descriptions linguistiques en la faisant considérer comme accessoire, voire paralinguistique. Or, du point de vue de la forme, c'est bien la prosodie qui entretient le rapport de proximité maximale avec la musique. Et sa réalisation acoustique n'entretient pas de rapport arbitraire avec son signifié. Il faudrait donc prendre la précaution de comparer ce qui est comparable et il est étonnant qu'un musicien ne se soit pas montré plus sensible à la musique de la langue. Par ailleurs, les notions de signifiant et de signifié reposent directement sur la distinction que Saussure établit entre langue et parole. Mais une distinction de ce type n'a jamais été établie pour la musique :

« dans la science de la musique, on n'a pas établi de distinction terminologique correspondante [à la distinction langue / parole pour le langage]. Dans cet article, nous utiliserons le terme d'exécution de la musique pour désigner la réalisation acoustique de la musique » (Carlson et al., 1989 : 551).

Il reste donc très difficile de s'affranchir de son domaine de spécialité : un linguiste aura tendance à imposer des modèles créés au départ pour la linguistique, sans respecter les spécificités de la musique.

Ainsi, celle-ci est souvent caricaturée à l'extrême et il est hasardeux de parler de la musique en général, alors qu'en réalité, on ne se réfère qu'à une culture musicale circonscrite géographiquement et historiquement : il devient facile de généraliser et de conclure que les limitations propres à cette musique sont celles de la musique en général.

3.4. Capacité cognitive commune

Dans ce contexte, une approche comme celle de Lerdahl et Jackendoff (1983) se révèle intéressante car ces chercheurs essaient justement de respecter les propriétés de la langue et celles de la musique sans caricaturer l'une en fonction de l'autre. Ils reconnaissent les similitudes structurelles entre la langue et la musique, notamment entre la structure prosodique de la langue et la structure des trames temporelles en musique, mais ne veulent pas les assimiler l'une à l'autre. Ils font l'hypothèse que la langue et la musique sont des

« manifestations différentes d'une capacité cognitive plus générale et abstraite, à savoir la segmentation et l'organisation temporelles, qui caractérisent aussi d'autres activités, comme la danse » (Clarke, 1989 : 31).

Dans ce sens, langue et musique seraient *« deux caractéristiques analogues, ayant une certaine forme d'origine cognitive commune »*, qui se seraient développées *« dans leur domaine respectif, selon les principes spécifiques du domaine »* (Clarke, 1989 : 31). Cette hypothèse rejoint d'une certaine manière celle de Condillac, Rousseau et Assafjew sur la spécialisation de chacun des deux systèmes, spécialisation qui serait responsable de leur différenciation au cours de l'histoire de l'humanité (cf. partie 1.4.). Cette perspective d'une origine cognitive commune de la langue et de la musique prend une dimension tout-à-fait captivante lorsque l'on se place non plus du point de vue de la phylogénèse, mais de l'ontogénèse. En effet, pendant la première année de la vie, langue et musique sont dans une certaine mesure indifférenciées (cf. Chapitre II). Elles pourraient donc dériver toutes deux de certaines propriétés fonctionnelles propres à notre système auditif nerveux (cf. Chapitre VII, 3. discussion).

Bibliographie chapitre I

- Awake ! (1997). Do african drums really talk ? Watchtower Bible and Tract Society of New-York, July 22, 21-23.
- Barraud, Henry (2002). La mélodie. Dans *Encyclopedia Universalis* en ligne.
- Beier, U. (1976). The talking drum of the Yoruba. In Umiker and Sebeok, 690-692.
- Benveniste, Emile (1951). *La Notion de « Rythme » dans son expression linguistique*. Paris : P.U.F., 402 p.
- Berio, Luciano (1998). *Sequenzas*. Livret d'accompagnement du CD, Deutsche Grammophon (457 038-2). Ensemble Intercontemporain.
- Bigand, Emmanuel (1994). Contributions de la musique aux recherches sur la cognition auditive humaine. Dans Mac Adams et Bigand, 249-298.
- Bolinger, Dwight L. (1998). Intonation in American English. In Hirst and Di Cristo, 45-55.
- Boulez, Pierre (1963). *Dire, Jouer, Chanter. La Musique et ses Problèmes Contemporains*. Paris : Juillard.
- Brelet, Gisèle (1949). *Le Temps Musical : Essai d'une Esthétique Nouvelle de la Musique*. Paris : P.U.F., 842 p.
- Burnet, James (1776). Réédité en 1974. *Of the Origine and Progress of Language*. New-York: Olms, vol. III.
- Busnel, René Guy and Classe, André (1976). *Whistled Language*. New-York: Springer, 117 p.
- Carlson, Rolf, Friberg, Anders, Frydén, Lars, Granström, Björn et Sundberg, Johan (1989). La parole et l'exécution de la musique : parallèles et contrastes. Dans Mac Adams et Deliège, 551-571.
- Chion, Michel (1983). *Guide des Objets Sonores : Pierre Schaeffer et la Recherche Musicale*. Paris : Buchet-Chasnel, 187 p.
- Clarke, F. E. (1989). Considérations sur le langage et la musique. Dans Mac Adams et Deliège, 23-43.
- Cohen, Antonie and Nootboom, Sieb G. (1975). *Structure and Process in Speech Perception*. Berlin: Springer, 351 p.
- Condillac (1746). *Essai sur l'Origine des Connaissances Humaines*. Paris : Armand Colin, ed. de 1924, 225 p.
- Cooke, Deryck (1959). *The Language of Music*. Londres: Oxford University Press.
- Cooper, G. and Meyer, L. B. (1960). *The Rhythmic Structure of Music*. University of Chicago Press.
- Cowan, Georges M. (1948). Mazateco whistled speech. In *Language*, 24/3, 280-286.
- Crystal, David (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press, 381 p.
- Crystal, David (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press, 472 p.
- Dauer, Rebecca M. (1983). Stress-timing and syllable-timing reanalysed. In *Journal of Phonetics*, 11, 51-62.
- Delattre, Pierre (1969). L'intonation par les oppositions. Dans *Le Français dans le Monde*, 64, 6-13.
- Delcloque, Philippe (1999). *Speech Technology Applications in CALL*. Dundee: University of Albertay Dundee, 87 p.
- Di-Cristo, Albert (1975). *Soixante et Dix Ans de Recherches en Prosodie : Bibliographie Alphabétique, Thématique et Chronologique*. Aix-en-Provence : Editions de l'Université de Provence, Travaux de l'Institut de Phonétique d'Aix-en-Provence, 351 p.
- Di-Cristo, Albert (1998). Intonation in French. In Hirst et Di Cristo, 195-218.
- Di Cristo, Albert (1999). Le cadre accentuel du français contemporain : essai de modélisation. Dans *Langues*, 3/2, 184-205.
- Didier, Béatrice (1982). Condillac et la Musique. Dans Sgard, 395-405.

- Dodane, Christelle, Konopczynski, Gabrielle, Santi, Serge et Espesser Robert (1999). Fo declination line: more evidence from infant's and children's speech (9-48 months). *Actes du Colloque ICPH's*, San Francisco, 2161-2164.
- Dom de Malherbe (1946). *Le Chant Grégorien*. Paris : Lyra Dei.
- Dom Gajard, Joseph (1951). *La Méthode de Solesmes : ses Principes Constitutifs, ses Règles Pratiques d'Interprétation*. Paris : Desclée et Cie, 94 p.
- Dowling, Jay W. and Harwood, Dane L. (1986). *Music Cognition*. Orlando: Academic Press, 258 p.
- Drake, Carolyn (2002). Temporal organisation of sound sequences: Same processes in music, environmental scenes and speech ? ISCA Workshop "Temporal Integration in the Perception of Speech", 8-10 avril, Aix en Provence.
- Drake, Carolyn, Penel, A. and Bigand, Emmanuel (2000a). Tapping in time with mechanically and expressively performed music. In *Music Perception*, 18, 1-24.
- Drake, Carolyn, Jones, Riess, Mari and Baruch, Clarisse (2000b). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. In *Cognition*, 77, 251-288.
- Encyclopædia Universalis (2002). *Base de données Universalis en ligne* (consultée en février 2002) : <http://www.universalis-edu.com>.
- Essien, Jimmy Akpan (2000). *Contribution à la Recherche sur la Perception des Tons du Yoruba : Evidences Expérimentales à partir des Tambours, des Signaux de la Parole et de la Synthèse*. Thèse de Doctorat, Université Paris III, non publiée, 426 p.
- Faure, Georges (1962). *Recherches sur les Caractères et le Rôle des Éléments Musicaux dans la Prononciation Anglaise (Essai de Description Phonologique)*. Thèse de Doctorat d'Etat. Paris : Didier, 380 p.
- Fonagy, Yvan and Magdics, Klara (1963). Emotional patterns in intonation and music. In *Zeitschrift für Phonetik, Sprach-Wissenschaft und Kommunikations-Forschung*, 16 (1-3), 293-326.
- Fonagy, Yvan, Guzman, Marina et Bérard, Eva (1976). Comment mesurer l'accent d'intonation ? Dans *Travaux de l'Institut d'Études Linguistiques et Phonétiques*, 2, 41-61.
- Fonagy, Yvan, Bérard, Eva et Fonagy, Judith (1983). Clichés mélodiques. Dans *Folia Linguistica*, XVII, 153-185.
- Fraisse, Paul (1974). *La Psychologie du Rythme*. Paris : PUF, 244 p.
- Fréron, Alain (2002). Matériau. *Encyclopedia Universalis* en ligne.
- Gagnard, Madeleine (1987). *La Voix dans la Musique Contemporaine et Extra-Européenne*. Van de Velde, 117 p.
- Gajewski, Borelas (1902). *Grammaire du Solrèsol ou Langue Universelle de François Sudre*. Paris, 44 p. (disponible à l'adresse suivante : <http://www.ptialaska.net/~srice/solresol/intro.htm>).
- Grosjean, François (1983). How long is the sentence ? Prediction and prosody in the on-line processing of language ? In *Linguistics*, 21, 501-529.
- Grosjean, François et Hirt, Cendrine (1996). Using prosody to predict the end of sentences in English and French: normal and brain-damaged subjects. In *Language and Cognitive Processes*, 11 (1-2), 107-134.
- Guaïtella, Isabelle (2000). Les systèmes de notation de la prosodie et du geste : pourquoi et comment les utiliser ? Dans *Guimbretière*, 165-184.
- Guimbretière, Elizabeth (2000). *Apprendre, Enseigner, Acquérir : la Prosodie au Cœur du Débat*. Rouen : Publications de l'Université de Rouen, 312 p.
- Guimbretière, Elisabeth (2000). Entretien avec Ivan Fonagy : Mes rencontres avec l'intonation. Dans *Guimbretière*, 151-163.
- Handel, Stephen (1989). *Listening: an Introduction to the Perception of Auditory Events*. Cambridge: MIT Press, 597 p.
- Hardcastle, William J. and Laver, John (1997). *The Handbook of Phonetic Sciences*. Padstow: Blackwell Publishers, 904 p.
- Herzog, G. (1945). Drum signalling in a West African Tribe. In *Word*, 1, 217-238.

- Hirsh-Pasek, Kathryn, Kemler, Nelson, Deborah, Jusczyk, Peter W., Wright, Cassidy, K., Druss, B. and Kennedy, L. (1987). Clauses are perceptual units for young infants. In *Cognition*, 26, 269-286.
- Hirst, Daniel (1998). Intonation in British English. In Hirst et Di Cristo, 56-78.
- Hirst, Daniel and Di Cristo, Albert (1998). *Intonation Systems: a Survey of Twenty Languages*. Cambridge: University Press, 487 p.
- Hirst, Daniel and Di Cristo, Albert (1998). A survey of intonation systems. Dans Hirst et Di Cristo, 1-44.
- Jackendoff, Ray S. (1991). Musical parsing and musical affect. In *Music Perception*, 9, 199-230.
- Jakobson, Roman et Waught, Linda, R. (1980). *La Charpente Phonique du Langage*. Paris : les Editions de Minuit, 336 p.
- Jespersen, Otto (1922). *Nature, Evolution et Origine du Langage* (Language: It's nature, development and origin). Paris : Payot, 436 p. Première édition, George Allen and Unwin, Ltd, London.
- Jones, Daniel (1909). *Intonation Curves*.
- Jones, Riess Marie et Yee, William (1994). L'attention aux événements auditifs : le rôle de l'organisation temporelle. Dans Mc Adams et Deliège, 75-121.
- Jones, Riess Marie and Boltz, M. (1989). Dynamic attending and responses to time. In *Psychological Review*, 96, 459-491.
- Jusczyk, Peter W. (1989). Perception of cues to clausal units in native and non-native languages. Paper presented at the *Biennial Meeting of the Society for Research in Child Development*, Kansas City, Mo., April.
- Kaiser, L. (1957). *Manual of Phonetics*. Amsterdam: North-Holland publishing company, 460 p.
- Klatt, Dennis K. and Cooper, William E. (1975). Perception of segment duration in sentence contexts. In Cohen and Nooteboom, 69-86.
- Konopczynski, Gabrielle (1978). Un essai original de transcription musicale de la prosodie. Wilhelm Wundt : *Völkerpsychologie (1900-1912)*. Dans *Travaux de l'Institut de Phonétique de Strasbourg*, 10, 82-97.
- Krumhansl, Carol L. (1983). Perceptual structures for tonal music. In *Music Perception*, 1, 28-62.
- Krumhansl, Carol L. Shepard, R. N.(1979). Quantification of the hierarchy of tonal functions in music. In *Journal of Experimental Psychology: Human Perception and Performance*, 5, 579-594.
- Krumhansl, Carol L. and Jusczyk, Peter W. (1990). Infant's perception of phrase structure in music. In *Psychological Science*, 1, 70-73.
- Lacheret-Dujour, Anne et Beaugendre, Frédéric (1999). *La Prosodie du Français*. Paris : C.N.R.S. Langage, 353 p.
- Leipp, Emile (1971). *Acoustique et Musique*. Paris : Masson. 3^{ème} éd. (1980), 352 p.
- Léon, Pierre et Léon, Monique (1964). *Introduction à la Phonétique Corrective : à l'Usage des Professeurs de Français à l'Étranger*. Paris : Hachette, 98 p.
- Léon, Pierre et Martin, Philippe (1969). *Prolégomènes à l'Étude des Structures Intonatives*. Dans *Studia Phonetica*, n°2. Paris : Didier, 225 p.
- Léon, Pierre (1992). *Phonétisme et Prononciations du Français*. Paris : Nathan Université, 192 p.
- Lerdahl, Fred (1985). Théorie générative de la musique et composition musicale. Quoi ? Quand ? Comment ? Dans *La Recherche Musicale*. Paris : Bourgeois/IRCAM, 101-120.
- Lerdahl, Fred and Jackendoff, Ray S. (1985). *A Generative Theory of Tonal Music*. Cambridge: M.I.T. Press, 368 p.
- Lieberman, Philip (1963). Some effects of semantic and grammatical context on the production and perception of speech. In *Language and Speech*, 6, 172-187.
- Lindblom, Björn and Sundberg, J. (1970). Towards a generative theory of melody. In *Svensk Tidskrift för Musikforskning, Swedish Journal of Musicology*, 52, 171-881.
- Lipsitt, Lewis P. and Rovee-Collier, Carolyn K. (1981). *Advances in Infancy Research*. Vol I : XXII. Norwood: Ablex, 267 p.

- Mac Adams, Stéphane et Deliège, Irène (1989). *La Musique et les Sciences Cognitives*. Liège : Pierre Mardaga, 649 p.
- Mac Adams, Stéphane et Bigand, Emmanuel (1994). *Penser les Sons : Psychologie Cognitive de l'Audition*. Paris : PUF, 402 p.
- Marler, Peter R. (1969). Animals and Man: Communication and its Development. In Roslansky, 25-62.
- Marler, Peter R. (1977). The evolution of communication. In Sebeok, 45-70.
- Massin, Jean et Massin, Brigitte (1983). *Histoire de la Musique Occidentale*. Messidor-Temps Actuels. Ed. utilisée, 1998. Fayard, 1312 p.
- Meigret, Louis (1550). *Le Tretté de la Gramèze Françoeze*. Nach der einzigen Pariser Ausgabe. Herausgegeben von Wendelin Foerster, Heilbronn, 1988.
- Menezes, Flo (1993). *Luciano Berio et la Phonologie : une Approche Jakobsonienne de son Oeuvre*. Berlin : Peter Lang, Publications Universitaires Européennes, 278 p.
- Mettas, Odette (1971). *Les Techniques de la Phonétique Instrumentale et l'Intonation*. Paris : P.U.F., 166 p.
- Michon, John A. (1967). *Timing in Temporal Tracking*. Soesterberg: Institute for Perception RVO-TNO, 127 p.
- Morgan, James L., Meier, R. P. and Newport, Elissa L. (1987). Structural packaging in the input to language learning: contributions of prosodic and morphological marking to the acquisition of language. In *Cognitive Psychology*, 19, 498-550.
- Nazzi, Thierry, Kemler Nelson, Deborah, Jusczyk, Peter and Jusczyk, A.M. (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. In *Infancy*, 1/1, 123-147.
- Osmond-Smith, David (1989). Entre la musique et le langage : vue depuis le pont. La parole et l'exécution de la musique : parallèles et contrastes. Dans Mac Adams et Deliège, 137-148.
- Palmer, C. and Krumhansl, Carol L. (1990). Mental representation for musical meter. In *Journal of Experimental Psychology: Human Perception and Performance*, 16, 728-741.
- Papousek, Mechthild and Papousek, Hanus (1981). Musical elements in the infants vocalizations: Their significance for communication, cognition and creativity. In Lipsitt, 163-224.
- Passy, Paul (1890). *Etudes sur les Changements Phonétiques et leurs Caractères Généraux*. Paris : Firmin-Didot, 270 p.
- Patel, A. and al. (1998) Processing prosodic and musical patterns: A neuropsychological investigation. In *Brain Language*, 61, 123-144.
- Perrot, Jean (1997). *Polyphonie pour Yvan Fonagy*. Paris : L'Harmattan, 530 p.
- Pierrehumbert, Janet (1980). *The Phonology and Phonetics of English Intonation*. PhD thesis, MIT, Indiana University Linguistic Club.
- Pike, Kenneth L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press, 203 p.
- Povel, D. J. (1981). Internal representations of simple temporal patterns. In *Journal of Experimental Psychology: Human Perception and Performance*, 7, 3-18.
- Rivière-Raverlat, Jacqueline (1997). *Développer les Capacités d'Ecoute à l'Ecole*. Paris : PUF, 207 p.
- Roslansky, J. D. (1969). *Communication. A Discussion at the Nobel Conference*. Amsterdam: North Holland Publishing Company.
- Rossi, Mario (1999). *L'Intonation, le Système du Français : Description et Modélisation*. Paris : Ophrys, 237 p.
- Roudet, Léonce (1899). Méthode expérimentale pour l'étude de l'accent. Dans *La Parole*, 5, 321-344.
- Roudet, Léonce (1910). *Eléments de Phonétique Générale*. Paris : Welter, 363 p.
- Rousseau, Jean-Jacques (1763). *Essai sur l'Origine des Langues où il est parlé de la Mélodie et de l'Imitation Musicale*. Ed. de 1990. Paris : Gallimard, 284 p.
- Rousselot, l'abbé (1901-1908). *Principes de Phonétique Expérimentale*. Paris : Welter, 2ème vol., 1252 p.
- Ruwet, Nicolas (1972). *Langage, Musique, Poésie*. Paris : Seuil, 247 p.
- Sauvanet, Pierre (2002). Le rythme. Dans *Encyclopedia Universalis* en ligne.

- Schaeffer, Pierre (1966). *Traité des Objets Musicaux*. Paris : Seuil, 712 p.
- Schaffer, L. H. (1982). Rhythm and timing in skill. In *Psychological Review*, 83, 109-122.
- Schuyesmans (2002). *Le Chant Grégorien*. Site Internet disponible à l'adresse suivante : <http://www.schuyesmans.be/gregoriaans/>
- Scripture, Edward, W. (1902). *The Elements of Experimental Phonetics*. Londre: Edward Arnold. Présente édition (1973), New-York, AMS Press Inc, 627 p.
- Sebeok, Thomas (1977). *How Animals Communicate*. London: Indiana University Press.
- Sgard, Jean (1982). *Condillac et les Problèmes du Langage*. Paris : Slatkine, 462 p.
- Shadle, Christine H. (1997). The aerodynamics of speech. In Hardcastle and Laver, 33-64.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. and Hirschberg, J. (1992). ToBI: A standard for labelling English prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing*, 867-870.
- Sloboda, J. A. (1976). The effect of item position on the likelihood of identification by influence in prose reading and music reading. In *Canadian Journal of Psychology*, 30, 228-236.
- Soulez, Antonia (1997). L'essence de la phônè (à propos du Cratyle 426C-427E). Dans Perrot, 411-423.
- Steele, Joshua (1775). *Prosodia Rationalis or an Essay towards Establishing the Melody and Measure of Speech to be Expressed and Perpetuated by Peculiar Symbols*. London: Nichols, J.
- Todd, N. (1985). A model of expressive timing in tonal music. In *Music Perception*, 3, 33-57.
- Umiker, Jean and Sebeok, Thomas A. (1976). *Speech Surrogates*. Mouton. The Hague.
- Van Noorden, L. P. A. S. (1975). *Temporal Coherence in the Perception of Tone Sequences*. PhD. The Netherlands, Eindhoven University of Technology.
- Van Waesberghe, J. Smits (1957). Phonetics in its relation to musicology. In Kaiser, 372-384.
- Viret, Jacques (1987). *La Modalité Grégorienne : un Langage pour quel Message ?* Lyon : A Cœur Joie, 167 p.
- Vos, P. and Troost, J. (1989) Ascending and descending melodic intervals: statistical findings and their perceptual relevance. In *Music Perception*, 6, 383-396.
- Wing, A. M. and Kristofferson, A. B. (1973a). The timing of inter-response intervals. In *Perception and Psychophysics*, 13, 455-460.
- Wing, A. M. et Kristofferson, A. B. (1973b). Response delays and the timing of discrete motor responses. In *Perception and Psychophysics*, 14, 5-12.
- Wundt, Wilhelm (1900). *Völkerpsychologie. Eine Untersuchung der Entwicklungsgesetze von Sprache, Mythos und Sitte*. Leipzig : Engelmann, W., 3ème édition, 1912.
- Zwanenburg, Wiecher (1965). *Recherches sur la Prosodie de la Phrase Française*. Leiden : Universitaires Pers., 135 p.