

L'hétérophonie contagieuse : une nouvelle théorie des origines de la musique

Steven Brown

Si les discussions relatives aux origines de la musique étaient monnaie courante aux XVIII^e et XIX^e siècles (Condillac, 1746 ; Rousseau, 1781 ; Spencer, 1857, 1890 ; Darwin, 1871 ; 1872), le sujet n'a pas semblé éveiller beaucoup d'intérêt au XX^e siècle. Pourtant, il semble qu'un intérêt pour l'évolution de la musique soit réapparu à l'aube du XXI^e siècle (Wallin, Merker & Brown, 2000 ; Brown, 2000b ; Miller, 2000b ; Huron, 2001 ; Cross, 2001 ; Morley, 2002 ; Hauser & McDermott, 2003 ; Hagen & Bryant, 2003). Dans le présent article, je présenterai une nouvelle hypothèse relative à la manière dont la musique s'est développée. Les premières parties de l'article présentent une proposition que je fais de l'existence d'un précurseur, dans l'évolution, à la fois du chant et du langage, basé sur un processus vocal auquel je donne le nom d'« hétérophonie contagieuse ». La dernière partie, quant à elle, décrit la manière dont ce précurseur peut avoir donné naissance, chez l'homme, aux arrangements uniques des discours langagiers et musicaux. Conformément à cette hypothèse, je décris la distinction entre « intégration » et « alternance » dans les évolutions respectives du discours musical et linguistique. Toutefois, avant de présenter en détail cette hypothèse, je ferai une brève incursion dans les théories historiques relatives à l'évolution de la musique, cela afin de fournir un contexte historique aux idées à élaborer.

DEUX THÉORIES DE L'HARMONIE : « THÉORIE SONORE » ET « THÉORIE DU SON »

Lorsqu'on aborde les théories acoustiques des origines de la musique en général (et de l'harmonie

en particulier), on peut distinguer deux classes de modèles. J'y ferai référence en utilisant les termes « sonore » (« sound ») et « du son » (« sounding »). On peut faire remonter la théorie « sonore » aux travaux de Jean-Philippe Rameau (par exemple, 1722) et à sa théorie classique du *corps sonore*. Rameau a avancé une théorie qui, en son temps, a été considérée dans le monde de la théorie musicale comme une révolution, au moins au même titre que les lois de la physique énoncées par Newton. Rameau considérerait que lorsque des objets se mettent à vibrer (comme par exemple lorsqu'on pince une corde de violon), non seulement il y a vibration principale de l'objet à une fréquence que nous percevons phénoménologiquement en tant que hauteur, mais, par rapport à cette vibration principale, il se produit simultanément toute une série de vibrations (appelées harmoniques). Ces harmoniques se produisent à des fréquences qui sont des multiples entiers de la fréquence fondamentale. Selon Rameau, ce sont donc ces harmoniques qui constituent la source de l'harmonie. Il considérerait cette dernière comme une caractéristique intrinsèque du son lui-même, comme la caractéristique première de la musique, dont toutes les autres provenaient. Depuis Helmholtz, les théories psychoacoustiques et physiologiques ont considéré les séries d'harmoniques comme le fondement physique des origines des intervalles musicaux et de la perception de la consonance en musique (*cf.* par exemple, Tramo, Cariani, Delgutte & Braidà, 2001).

Malgré son attrait réductionniste, cette vision des choses pose quelques problèmes. La théorie du *corps sonore* constitue une théorie acoustique qui s'applique à tous les objets physiques ; la spécificité

musicale en est donc limitée. Tout objet (par exemple, un verre), lorsqu'on le fait vibrer, génèrera une série de fréquences sonores qui ne sont pas très différentes de celles d'un instrument de musique conventionnel (et d'ailleurs le glass harmonica, ou hydrocristalophone, se base sur ce principe). Toutefois, à un niveau plus fondamental, je pense qu'il ne faut pas chercher les origines de la musique dans des mécanismes de *perception* sonore, mais plutôt dans des mécanismes de *génération* sonore ; c'est à ce niveau que les théories « du son » entrent en jeu. On peut considérer que les racines de la théorie « du son » se trouvent dans les idées avancées par Jean-Jacques Rousseau (par exemple, 1781) au sujet de l'importance de la mélodie dans l'évolution de la musique. Pour Rousseau, la voix et la mélodie sont les caractéristiques les plus essentielles de la musique (Thomas, 1995). Il considère qu'il n'existe aucun lien plus fort que celui qui unit la musique et la voix, et c'est ce qui, par essence, fait de la musique une activité de *communication*, contrairement au tintement du verre même s'il résonne magnifiquement de toutes ses harmoniques. Pour Rousseau, la mélodie, en musique, trouve ses origines non seulement dans le son, mais également dans la communication, et plus précisément dans l'expression vocale de l'émotion : « La mélodie en imitant les inflexions de la voix exprime la plainte, les cris de douleur ou de joie, les menaces, les gémissements ; tous les signes vocaux des passions sont de son ressort. » (Rousseau, 1781/1990, p. 123-124). Ceci dit, Rousseau ne s'est pas intéressé à la question primordiale de savoir pourquoi, chez l'homme, la musique comporte autant de mélange acoustique (c'est-à-dire autant d'harmonies), car il semblait trop occupé à réfuter la théorie de Rameau, selon laquelle c'est l'harmonie qui constitue l'élément fondamental de la musique (cf. Thomas, 1995, pour une discussion de la rivalité existant entre Rameau et Rousseau).

De l'harmonie, je propose une théorie « du son » qui va dans le sens de celle de Rousseau par rapport à son attachement à l'expression vocale, mais qui dépasse l'accent exclusif qu'il met sur la

vocalisation soliste. En ce sens, l'harmonie ne proviendrait pas en soi des harmoniques produites par les corps en vibration, mais plutôt du *mélange des voix*, donc du mélange choral. En d'autres termes, c'est l'*harmonisation* qui se trouve à la base de l'harmonie. Fidèle aux idées avancées par Rousseau en ce qui concerne la mélodie, la théorie harmonique « du son » est une théorie vocale qui s'applique exclusivement aux sons produits par les animaux lorsqu'ils communiquent. Elle met l'accent sur les processus de production plutôt qu'uniquement sur les processus perceptuels. Dans le présent article, j'utiliserai les termes « mélange » et « chœur » par référence aux vocalisations collectives émises dans un but de coopération par des espèces vivant en groupes ou en couples. C'est à l'explication de cette capacité de mélange coopératif intentionnel que les théories de l'évolution musicale doivent s'atteler. Le mélange choral, à mon sens, ne fait donc pas référence à une situation de simple chevauchement ou simultanéité, mais bien à une situation dans laquelle des espèces animales vivant en groupes ou en couples vocalisent ensemble, avec une intention commune de communication, en général dans une optique de protection de leur territoire. Si on accepte cette définition, le chœur des oiseaux qu'on entend à l'aube (des oiseaux d'espèces différentes qui chantent simultanément et de multiples mâles d'une même espèce qui sont en compétition les uns avec les autres pour être entendus par des partenaires potentiels) ne constitue pas un exemple de mélange choral : il s'agit simplement d'un exemple de simultanéité. De même, un ensemble de plusieurs centaines de traders à la Bourse de New York en train de crier leurs offres en même temps ne peut raisonnablement être considéré comme un exemple de chœur. C'est l'usage non discriminant qu'a fait l'éthologie du concept de « chœur » qui se trouve à l'origine d'un certain nombre de problèmes rencontrés par les chercheurs qui s'intéressent aux origines de la musique. Dans le présent article, j'utiliserai uniquement les termes « chœur » et « mélange » lorsque je ferai référence aux processus groupaux de mélange collectifs et coopératifs qui surviennent dans les

chœurs. La simultanéité (et même la synchronisation) ne se révèle pas suffisante.

Un des principaux arguments utilisés afin de donner la priorité à la production par rapport à la perception dans l'évolution de la communication musicale chez l'être humain est l'observation selon laquelle les organes/voies auditifs/-ives des humains et des chimpanzés présentent de fortes similarités mais, que, par contre, leurs voies vocales sont très différentes; ce qui donne, d'un côté, une espèce capable de chanter et de produire un discours articulé et, d'un autre côté, une autre espèce incapable de le faire. Si on a souvent cité des modifications périphériques (comme par exemple la descente laryngée) comme facteurs critiques de cette évolution (Lieberman, Klatt & Wilson, 1969), on reconnaît maintenant que des modifications du système nerveux revêtent une importance au moins égale (Fitch, 2000). Par exemple, le cortex moteur primaire constitue un médiateur critique de la vocalisation volontaire chez l'être humain. Les lésions bilatérales de ce cortex moteur primaire ou de la voie corticobulbaire associée mènent à des troubles moteurs débilants de la parole, connus sous les noms génériques de paralysie pseudobulbaire ou de dysarthrie spastique (Murphy, 2005). De plus, des lésions transitoires du cortex moteur par le biais d'une stimulation magnétique transcranienne sont susceptibles de mener à de graves troubles du langage, et même à son arrêt complet (Epstein, Meador, Loring, Wright, Weissmann, Sheppard, Lah, Puhlovich, Gaitan & Levy, 1999; Stewart, Walsh, Frith & Rothwell, 2001). Chez les primates non humains, on ne dispose pas de preuves établissant que le cortex moteur primaire soit même impliqué dans la vocalisation, dans la mesure où, ici, les lésions n'affectent pas la production des appels (Kirzinger & Jürgens, 1982). De plus, si on soupçonne les humains de disposer d'une projection neurale directe du cortex moteur primaire vers le noyau ambigu (centre principal du tronc cérébral, responsable du contrôle de la phonation), ce n'est pas le cas des primates non humains (Jürgens, 1992; Simonyan & Jürgens, 1993). Chez les humains, le

cortex auditif associatif du lobe temporal supérieur est relié par un faisceau principal aux centres de planification vocale du lobe frontal inférieur (Catani, Howard, Pajevic & Jones, 2002), ce qui crée les fondements neuraux de l'imitation vocale. On débat encore actuellement des détails de cette connectivité chez le singe et il n'est même pas encore certain que les aires préfrontales ventrales qui reçoivent l'input auditif soient des aires vocales (*cf.* toutefois, Petrides & Pandya, 2001). Le résultat comportemental principal de ces différences inter-espèces est qu'aucun primate, en dehors de l'humain, n'est capable de l'apprentissage vocal phonatoire (Egnor & Hauser, 2004) sous-tendant l'acquisition du chant et du langage par l'imitation vocale. L'observation du phénomène de « convergence des appels » chez les chimpanzés mâles (Mitani & Gros-Louis, 1998) – dans lequel les appels en chœur d'animaux entretenant les liens sociaux les plus forts convergent en une structure acoustique – révèle une sorte de plasticité vocale adulte qui pourrait bien constituer un précurseur phylogénétique de l'imitation vocale chez l'être humain.

Une conséquence de la théorie sonore de l'harmonie est constituée par l'idée selon laquelle la capacité de percevoir et de produire l'harmonie est essentiellement circonscrite à un système mélodique plutôt qu'à un module harmonique spécifique et dissociable. Je pose comme hypothèse que les capacités humaines de mélodie et d'harmonie ont co-évolué en tant que système neural unique (*cf.* Brown, Parsons, Martinez, Hodges & Fox, 2004). Cette vision des choses va dans le sens de Rousseau car elle implique que l'harmonie provient d'un système mélodique de base (ou qu'elle y est circonscrite) et, de plus, vocal. Ceci est attesté par le fait que, dans les diverses musiques du monde, la forme la plus répandue de mélange de hauteurs n'est pas l'homophonie (c'est-à-dire la texture musicale dans laquelle toutes les voix ont des mouvements plus ou moins semblables, comme dans la musique occidentale, faite d'accords), mais plutôt la polyphonie multi-couches, dans laquelle des motifs mélodiques courts sont tissés, entrelacés. Les exemples les plus

documentés en sont peut-être les ensembles instrumentaux et chœurs polyphoniques complexes des Pygmées d'Afrique centrale (Arom, 1991). Il n'est donc pas déraisonnable de spéculer sur le fait que le système harmonique chez l'être humain provient d'un système mélodique de base dans lequel des parties individuelles ont été temporairement mélangées avec d'autres à la suite d'évolutions du traitement temporel (*cf. infra*, la discussion au sujet de l'«intégration verticale»). Si elle devait se révéler exacte, cette hypothèse bouleverserait la théorie traditionnelle de la progression évolutionnaire telle que formulée par Curt Sachs (1943); cela voudrait dire que la polyphonie (ou, en tout cas, l'hétérophonie) a, dans l'histoire de l'être humain, précédé la monophonie (texture musicale dans laquelle toutes les voix chantent la même partie).

HÉTÉROPHONIE CONTAGIEUSE

En gardant à l'esprit notre vision de la «théorie du son» de l'évolution musicale, j'aimerais présenter au lecteur un précurseur possible dans l'évolution de la musique et de la parole chez l'homme, un précurseur qui présente de grandes similitudes avec le comportement d'appel qu'on retrouve chez bon nombre d'espèces animales contemporaines. Le précurseur en question est constitué par un système de mélange vocal basé sur ce que j'appelle «l'hétérophonie contagieuse». Je définis ce terme comme suit: une vocalisation en groupe dans laquelle chaque individu produit une variation par rapport à un type d'appel similaire, mais dans laquelle chaque membre du groupe réalise un appel non synchrone; la vocalisation du groupe tout entier se produit au terme d'un processus séquentiel de dissémination, de contagion. Un exemple de ce phénomène est constitué par les hurlements des loups. Chaque loup lance un appel similaire, mais le chœur résultant est temporellement mal mixé, créant ainsi un type d'hétérophonie. On peut, dans la nature, identifier bon nombre d'autres exemples de ce type de chœur: les cris des groupes de singes hurleurs ou, chez le chimpanzé, le chœur des «pant hoots»,

le rugissement des lions ainsi que le style de chant hétérophonique des baleines à bosse. L'autre caractéristique importante de cette hypothèse, c'est la contagion. Il est effectivement intéressant de constater que la contagion se retrouve en bonne place dans tous les exemples de vocalisation hétérophonique que nous venons de mentionner. Dès qu'un animal lance son appel, d'autres membres du groupe se joignent à lui par un processus de dissémination. On doit à Harrington (1989) une analyse du processus de hurlements en chœur par des meutes de loups en liberté. Ces hurlements constituent un processus de contagion: en effet, l'auteur s'est aperçu que le chœur débute souvent par les hurlements d'un seul loup, l'arrivée des autres loups dans le chœur «s'accélère plutôt qu'elle ne survient à un rythme relativement constant» (Harrington, 1989, p. 124). Dans la Figure 1a, on verra que l'intervalle de temps correspondant aux entrées successives des loups dans le chœur est de plus en plus court. Le résultat final de l'addition successive de parties à ce chœur est une augmentation de la texture ainsi qu'un remplissage de l'espace acoustique, ce qui apparaît clairement dans le sonagramme continu présenté dans la Figure 1b. Comme Harrington (1989) l'a bien montré, les hurlements deviennent plus courts, plus aigus et de fréquences plus modulées au fur et à mesure de la progression du chœur. On assiste à une forte augmentation de la variation spectrale au fur et à mesure de l'exécution du chœur, qui devient une sorte de «jam session».

On peut considérer que le concept d'hétérophonie contagieuse facilite grandement notre compréhension de la nature de la musique. En premier lieu, un système basé sur l'hétérophonie contagieuse place clairement la musique depuis son origine dans le domaine de la vocalisation en groupe. Ce qui correspond bien aux conclusions de plus d'un siècle de recherche ethnomusicologique, qui démontrent que la musique est, dans son sens le plus basique, un moyen de communication de groupe (Merriam, 1966; Dissanayake, 2000b), un moyen utile de consolidation des groupes sociaux (Brown, 2000b; Hagen & Bryant, 2003). De même, on pense que la

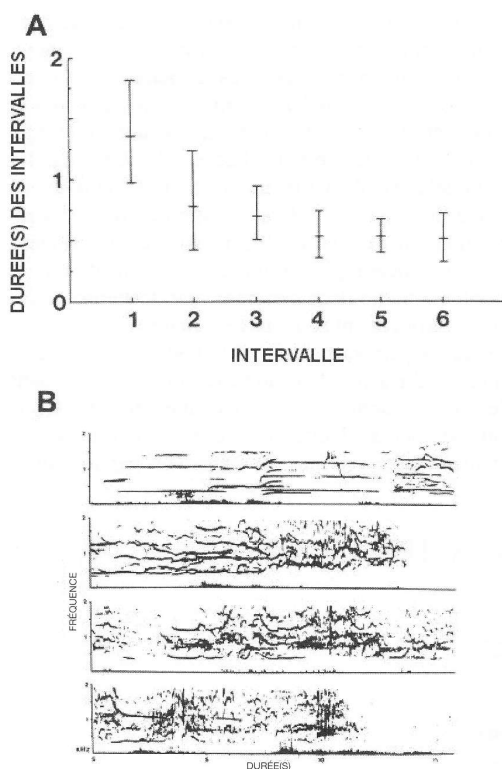


Figure 1.

Chœur de hurlements chez les loups

A. Le graphique présente les moments successifs d'entrée de loups individuels au début de chaque chœur. La durée de l'intervalle est constituée par le temps (exprimé en secondes) qui s'écoule entre les démarrages successifs des hurlements dans le chœur. L'intervalle 1 représente la période séparant le premier hurlement de l'arrivée du deuxième hurlement dans le chœur, l'intervalle 2 représente la période séparant le deuxième hurlement de l'arrivée du troisième hurlement, etc.

B. Sonagramme continu d'un chœur de loups (de haut en bas et de gauche à droite). À noter : l'augmentation de la modulation de fréquence au fur et à mesure de la progression du chœur. Reproduit avec permission de l'auteur (Harrington, 1989).

plupart des formes animales d'hétérophonie contagieuse que nous venons de mentionner jouent des rôles importants dans la fonction du groupe, notamment en ce qui concerne le maintien du territoire, l'identité du groupe et la cohésion sociale. C'est particulièrement clair chez les espèces qui pratiquent le duo : le chœur fournit alors un moyen important de marquer les territoires de manière acoustique et de maintenir les liens du couple, par exemple chez le gibbon (Geissmann, 2000), l'indri (Pollock, 1986) ou encore la pie australienne (Brown & Farabaugh, 1991). En deuxième lieu, on peut identifier, dans ces appels, les origines des systèmes tonals basiques. Les hurlements des loups débutent généralement par un simple intervalle montant qui se termine par une note soutenue (généralement à la quinte ou à la sixte), suivi de plusieurs *portamenti* en descente chromatique. Bien qu'il ne présente pas la même complexité mélodique que le chant des baleines ou que certaines formes élaborées de chant chez les oiseaux, ce type d'appels fournit néanmoins des éléments de tonalité simple. En troisième lieu, nous pouvons déceler dans un tel système les rudiments d'un véritable mélange collectif, menant finalement à ce qui a pu devenir la capacité d'homophonie chez les êtres humains modernes. Aux moments de l'entrée accélérée de plusieurs loups dans le chœur formé par le groupe, et au moment où plusieurs animaux maintiennent simultanément les notes correspondant aux hurlements, ce chevauchement des diverses parties présente, dans le règne animal, ce qui est peut-être la meilleure analogie avec l'accord en musique. Il est intéressant de considérer un moment les conséquences territoriales de ce type de mélange coopératif complexe. Harrington (1989) évoque un type d'effet « Beau Geste » du chœur des loups : le mélange complexe qu'il présente peut provoquer une surestimation importante de la taille du groupe (en tout cas par l'humain), ce qui accentue la fonction territoriale du hurlement. En quatrième lieu, le fait que tous les membres d'un groupe en train d'exécuter ces vocalisations en hétérophonie contagieuse créent des appels très similaires signifie que la complexité acoustique peut naître de variations dans le mélange plutôt que de

variations dans les seuls appels individuels. En d'autres termes, les petits répertoires individuels ne vont pas imposer de contraintes dans la création de la complexité au niveau du groupe. Je subodore que ce point s'applique aussi bien à de nombreuses formes de musique chez les êtres humains. Enfin, le rôle de la contagion dans un tel système fournit non seulement un mécanisme qui permet à la participation du groupe de se produire, mais également une indication importante de la manière dont ce type de système peut opérer, que ce soit au niveau psychologique ou au niveau neural. La contagion est une sorte de comportement de résonance dans lequel un état commun d'activité du groupe est atteint par le matching ou l'imitation. Il suffit de penser à la manière dont, lors d'un concert, les gens se lèvent à l'occasion d'une standing ovation : au départ, il n'y

a souvent qu'une seule personne qui se lève, mais peu de temps après c'est toute la salle qui est debout, par le biais d'un processus contagieux d'imitation. Le résultat final, c'est que tous les gens qui se trouvent dans la salle ont amené une modification à leur état initial (d'assis, ils sont maintenant debout) ; ils ont maintenant atteint un nouveau stade de résonance. Plus en rapport avec la vocalisation, la récitation déclamatoire de slogans dans les meetings politiques, ou les chants que le public entonne dans certaines manifestations sportives, surviennent souvent de manière contagieuse : quelques personnes incitent le reste de l'assistance à se joindre à eux. Un autre exemple : souvent, dans les fêtes d'anniversaire, c'est une seule personne qui se met à chanter « Joyeux Anniversaire » et donne le ton, rejointe spontanément par les autres

INTEGRATION - ALTERNANCE

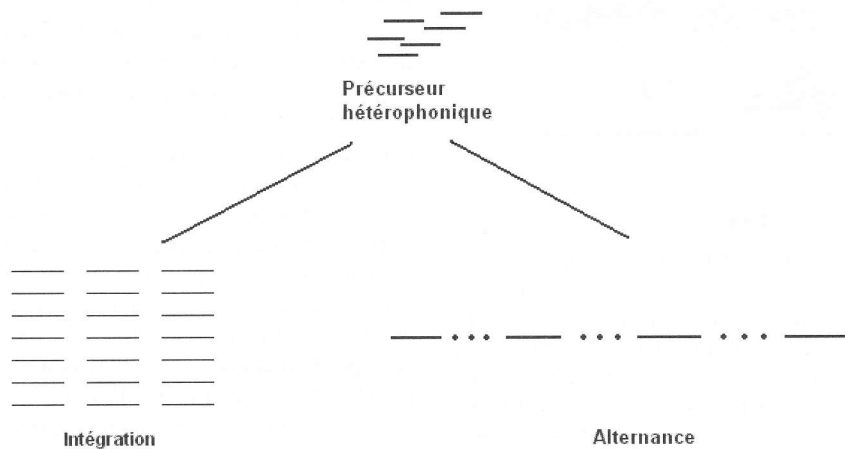


Figure 2.

Distinction entre intégration et alternance. Il y a progression évolutionnaire par laquelle un précurseur vocal basé sur l'hétérophonie contagieuse devient soit une forme intégrée de vocalisation (telle qu'on la retrouve souvent en musique) soit une forme alternante de vocalisation (telle qu'on la retrouve dans la plupart des formes de langage). En ce qui concerne l'intégration, je présente trois ensembles de parties intégrées séquencées afin de bien montrer l'organisation rythmique découpée dans le temps.

participants sur un ensemble donné de hauteurs. Et de ceux dont la voix n'est pas capable de s'aligner sur ces hauteurs on dit qu'ils n'ont pas l'oreille musicale.

Pour résumer, mon hypothèse est qu'un système vocal basé sur l'hétérophonie contagieuse constitue un précurseur raisonnable des systèmes qui deviendront, chez les êtres humains modernes, à la fois le langage et le chant. Il est intéressant de relever la théorie de Geissmann (2002), selon laquelle le duo a constitué, chez les gibbons, un trait ancestral et que le chant individuel s'est développé ensuite, par un abandon de ce duo, chez les espèces qui ne font plus actuellement que des appels individuels. De même, je pense que les appels en groupe peuvent avoir constitué un trait ancestral des humains. Mais comment la musique s'est-elle développée à partir du type de précurseur hétérophonique que j'ai suggéré? En d'autres termes, comment arriver, à partir d'un système de mélange non synchrone (comme par exemple les hurlements des loups), à un système d'assemblage très coordonné, comme celui qui caractérise par exemple la polyphonie musicale? Je fais la suggestion suivante : le chemin qui part de l'hétérophonie contagieuse pour arriver à la communication chez l'être humain s'est scindé en au moins deux sentiers complémentaires (cf. Figure 2) : celui 1) d'une *intégration* verticale, dont sont issus des styles d'exécution présentant des assemblages corrects bien ordonnés dans le temps (par exemple, la monophonie, l'homophonie ou la polyphonie) et celui 2) d'une *alternance* horizontale dans laquelle surviennent des processus de vocalisation bien fixés dans le temps mais alternants (ce qui est typique du langage chez l'être humain). Considérons l'exemple du chant monophonique, comme par exemple le « Joyeux Anniversaire » mentionné plus haut ; pour que, dans un groupe d'individus, on puisse parler de chant monophonique, au moins deux compétences sont requises : tout d'abord, la capacité de chanter à des hauteurs communes, et ensuite la capacité de s'aligner sur les entrées. Ceci met en lumière deux concepts connexes que je présenterai dans les para-

graphes suivants : l'*imitation vocale* et la *synchronisation métrique*.

L'IMITATION VOCALE : CORRESPONDANCE DE HAUTEURS

Bien que la plupart des vertébrés possèdent la capacité de vocaliser, très peu d'espèces sont capables d'imiter les sons vocalement. Les principales exceptions à cet état de choses sont : les chants des oiseaux, ceux des baleines à bosse, et ceux des êtres humains (Janik & Slater, 1997; 2000). Chez l'être humain, l'imitation vocale est importante, non seulement pour l'établissement de larges répertoires acoustiques flexibles dans le langage (par exemple, les phonèmes, les syllabes), comme dans la musique (par exemple, les ensembles de hauteurs), mais également pour la capacité d'imiter de manière précise et en temps réel les sons linguistiques, musicaux ou environnementaux, ce qui est important pour l'apprentissage de mélodies nouvelles. Il est maintenant établi que l'imitation vocale, au fur et à mesure du développement de l'être humain, fournit la base d'acquisition des sons communicationnels, comme les phonèmes et les schémas intonatifs pour le langage, et les hauteurs en matière de musique (Poulson, Kymissis, Reeve, Andreators & Reeve, 1991; Kuhl & Meltzoff, 1996; Papoušek, 1996; Studdert-Kennedy, 2000; Trehub, 2001). En d'autres termes, l'apprentissage vocal, chez l'être humain, est imitatif et dépendant d'interactions avec les modèles adultes lors des périodes critiques de son développement neurocognitif. Ce processus implique l'ontogenèse de voies neurales pouvant efficacement traduire les sons entendus lors de la communication en plans moteurs pour la vocalisation.

Si l'on ne connaît pas encore très bien le développement de ces voies neurales chez l'être humain, on a par contre constitué une importante base de données relative à la compréhension de processus analogues dans le cerveau des oiseaux chanteurs. Nous n'allons évidemment pas, dans le cadre de cet

article, passer en revue la littérature consacrée à la neurobiologie des oiseaux. Toutefois, le point principal à évoquer est celui de la distinction entre, d'un côté, un système neuromoteur nécessaire à la vocalisation elle-même et, d'un autre côté, un système séparé qui ne pilote pas directement la vocalisation, mais qui est nécessaire pour l'imitation vocale. Ces deux systèmes (qui, chez l'oiseau chanteur, sont référenciés respectivement comme les voies prosencéphaliques postérieure et antérieure) sont reliés l'un à l'autre mais peuvent, par l'effet de lésions, être fonctionnellement dissociés. Des lésions de la voie prosencéphalique postérieure qui surviennent après le développement du chant causent généralement la mutité (Simpson & Vicario, 1990), tandis que des lésions de la voie prosencéphalique antérieure lors de périodes critiques du développement peuvent effectivement empêcher un oiseau chanteur d'acquérir les modèles de chant spécifiques de son espèce sans affecter sa capacité générale de vocaliser (et d'ailleurs d'entendre également) (Scharff & Nottebohm, 1991 ; Brainard, 2004). Jarvis (2004) a fait remarquer que ce système double propre à l'oiseau présente de grandes similitudes avec une disposition propre au cerveau humain, dans lequel le cortex moteur primaire est en rapport avec la voie prosencéphalique postérieure, impliquée dans le contrôle vocal direct, tandis que le cortex prémoteur (par exemple, l'aire de Broca) est en rapport avec la voie prosencéphalique antérieure, qui, elle, est impliquée dans l'imitation vocale. Donc, que ce soit chez l'être humain ou chez l'oiseau chanteur, l'imitation vocale peut être modifiée par une partie du cerveau distincte de celle qui est impliquée dans le contrôle vocal direct. On peut supposer que ce type d'aire, contrairement à la voie motrice directe, devrait recevoir des inputs en provenance d'aïres auditives, comme cela semble établi pour l'aire de Broca chez l'être humain.

Au-delà de l'exigence d'imitation vocale au cours du développement du chant et de la parole, il faut que cette capacité puisse être maintenue tout au long du cycle de vie. Les êtres humains sont d'excellents imitateurs vocaux (Skoyles, 1998). Certains

en font d'ailleurs leur profession dans le show business. En situation d'expérimentation, des sujets sont capables de « coller » à une prose continue avec un retard d'à peine 70 millisecondes (Bailly, 2002 ; cf. également Marslen-Wilson, 1973 et Porter & Lubker, 1980 pour des rapports faisant état de latences de plus longues durées). Ce genre de latence est beaucoup plus court que n'importe quel type de temps de réaction dans les cas de réactions manuelles, ce qui suggère une connexion extrêmement directe entre l'audition et la production vocale. La capacité de l'être humain d'imiter vocalement peut, dans certaines conditions cliniques, devenir hyperactive. Parmi ces conditions, citons l'écholalie qu'on peut définir comme une « répétition involontaire du discours d'autrui » (Murphy, 2005, p. 359). Il s'agit là d'une caractéristique principale de l'autisme, du syndrome de Tourette, et d'autres troubles neurodégénératifs graves (Schuler, 1979). Dans les cas d'écholalie, l'imitation spontanée peut survenir dans un délai aussi court que 270 millisecondes à partir du démarrage de la production du stimulus (Fay & Coleman, 1977).

L'imitation joue également un rôle important dans les formes traditionnelles d'exécution musicale partout dans le monde. Deux exemples bien documentés sont d'une part le style de chant « déphasé » (*dulugu ganalan*, « son qui soulève par delà ») des Kalulis (peuple habitant les régions montagneuses du sud de la Papouasie Nouvelle Guinée) ainsi que les « jeux de gorge » (*Katajjaq*) des Inuits (Canada) dans lesquels un individu imite les sons produits par un autre, leurs bouches se touchant presque (Nattiez, 1999). En jazz ou en pop music, on retrouve des traditions de « rituels d'imitation » sur scène dans lesquels des musiciens réalisent à deux des improvisations alternées (instruments et timbres différents) : un musicien improvise une ligne mélodique et l'autre la répète (soit à l'identique, soit avec un phrasé légèrement différent, comme par exemple avec une ornementation ou une coda attachée). Bien que cela soit anecdotique, je voudrais également mentionner que le fait de « coller » aux voix joue un rôle important dans le

chant choral, où il y a une forte tendance à écouter les gens qui se trouvent placés près de vous et chantent dans la même voix que vous (dans mon cas, les barytons). Ceci peut se révéler bénéfique si vous avez des hésitations par rapport à certaines notes, mais également constituer un problème si votre voisin de pupitre commet des erreurs, car la tendance est alors de lui emboîter le pas. Enfin, l'amusie représente, quant à elle, une capacité amoindrie de faire correspondre la production vocale à une hauteur donnée. Tout comme bon nombre d'autres troubles développementaux et acquis touchant la voix, l'amusie est beaucoup plus fréquente chez l'homme que chez la femme (Howard & Angus, 1998). On a avancé l'idée qu'elle est congénitale (Ayotte, Peretz & Hyde, 2002; Peretz, Ayotte, Zatorre, Mehler, Ahad, Penhune & Jutras, 2002) et qu'elle est donc susceptible d'avoir une base génétique (Kalmus & Fry, 1980; Drayna *et al.*, 2001). D'autres interprétations sont néanmoins plausibles, par exemple la nécessité pour un enfant d'être exposé à de la musique et à de la pratique vocale lors d'une période critique de son développement cérébral. Il se pourrait très bien que la capacité de faire correspondre la voix à une hauteur donnée soit, un peu comme dans le cas de l'oreille absolue, une compétence qui disparaît si on ne l'utilise pas. Bien qu'à ma connaissance on ne dispose pas d'analyses ethnographiques publiées concernant l'amusie, je pense ne pas me tromper en disant qu'on la trouvera surtout chez les individus qui, lorsqu'ils étaient enfants, n'ont eu personne qui chantait pour eux, et chez ceux qu'on n'a pas encouragés à chanter.

On pense que l'existence des comportements d'imitation (comme la correspondance voix-hauteur) peut s'expliquer, notamment, par un système de « neurones miroirs ». Ces neurones miroirs constituent un type de cellules découvert pour la première fois chez le singe Rhésus, dans la région du cortex prémoteur qui pilote les mouvements de la main; on s'est aperçu qu'ils déchargeaient non seulement lors des mouvements de la main vers un objet, mais également lors de l'*observation* du

même type de mouvements de la main réalisés par un autre organisme. Un système miroir constitue essentiellement un système de matching observation/exécution (Rizzolatti, Fadiga, Gallese & Fogassi, 1996; Rizzolatti, Fadiga, Fogassi & Gallese, 1999). Dans les écrits que Rizzolatti *et al.* ont consacrés au sujet, le centre d'intérêt de ce type de système a surtout concerné le matching visuomaneuel. Je défends toutefois l'idée qu'un tel système constitue le fondement idéal de fonctions de matching *audiovocal* telles que le chant et le langage. Une caractéristique intéressante de la théorie des comportements de résonance par activation des neurones miroirs est le rôle attribué à l'aire de Broca dans le système de résonance. Cette aire de Broca fait partie du lobe frontal inférieur du cerveau et, depuis plus d'un siècle, on pense qu'elle joue un rôle primordial dans la production de la parole. Rizzolatti et Arbib (1998) ont fait remarquer que, chez le singe Rhésus, la région du cerveau dont ils ont démontré qu'elle présentait des neurones miroirs était l'homologue de l'aire de Broca chez l'homme. Ces chercheurs ont donc avancé l'hypothèse selon laquelle l'aire de Broca s'est développée à partir d'un système prémoteur impliqué dans la communication gestuelle. À l'appui de cette hypothèse, Iacoboni *et al.* (1999) ont démontré, lors d'une étude d'imagerie cérébrale, que l'opercule frontal (c'est-à-dire la partie inférieure de l'aire de Broca) était en activité lorsqu'il était demandé aux sujets d'observer, et ensuite d'imiter, des mouvements des doigts.

Mes collègues et moi-même avons mené des expériences d'imagerie cérébrale qui fournissent certains renseignements sur le rôle de cette aire du cerveau dans la production de chant imitatif (Brown *et al.*, 2004). Parmi les tâches de nature vocale que les sujets devaient accomplir au cours de cette étude, se trouvait une tâche de répétition mélodique: les sujets écoutaient une série de mélodies tonales monophoniques qu'ils ne connaissaient pas, et ils devaient rechanter chacune d'entre elles après qu'on l'ait jouée. Chaque mélodie durait six secondes, suivies de six autres secondes pour la

production de la réponse. Les résultats obtenus lors de cette étude ont clairement démontré que l'opercule frontal était activé bilatéralement pour l'exécution de cette tâche. (De nombreuses autres aires du cerveau étaient également activées, mais il s'agissait d'aires plus en rapport avec les processus fondamentaux de l'audition et de la vocalisation). Il est dès lors tentant de conclure que cette portion de l'aire de Broca est spécialisée dans les processus de *mise en conformité par rapport à des gabarits (template-matching)* qui sous-tendent l'imitation de la hauteur. Et, en fait, une autre caractéristique qui rend cette notion de *template-matching* importante, d'un point de vue neurobiologique, est constituée par l'activation parallèle de cette aire de planification vocale lors de l'exécution de tâches perceptuelles dans lesquelles il n'y a pas de vocalisation. Il est clair que le *template-matching* constitue un processus essentiel dans l'exécution de tâches de discrimination; un grand nombre d'études ont démontré qu'en matière musicale l'opercule frontal est actif lors des tâches perceptuelles de discrimination telles que la discrimination des hauteurs (Zatorre, Evans & Meyer, 1994; Griffiths, Johnsrude, Dean & Green, 1999; Zatorre & Binder, 2000), la discrimination des accords (Maess, Koelsch, Gunter & Friederici, 2001), la discrimination des durées (Griffiths *et al.*, 1999), la discrimination des intervalles de temps (Rao, Mayer & Harrington, 2001), celle des accords, tonalités et timbres (Koelsch, Gunter, Cramon, Zysset, Lohmann & Friederici, 2002) ou encore celle des erreurs en rapport avec la mélodie ou l'harmonie lors du déchiffrement de la partition (Parsons, 2001). Globalement, mon opinion est qu'une fonction de miroir dans l'aire de Broca est susceptible de nous éclairer autant au sujet des processus audiovocaux d'imitation qui sous-tendent la musique et le langage qu'à celui des processus de *matching* visuomanuel qui sous-tendent une possible origine gestuelle du langage (Rizzolatti & Arbib, 1998). Cela pourrait suggérer que le système du chant dans le cerveau humain s'est développé à partir d'un système de vocalisation basé sur l'imitation, éventuellement dérivé d'un précurseur hétérophonique.

Un avertissement s'impose ici. Au départ, l'activité des neurones miroirs a été invoquée afin d'expliquer « la compréhension de l'action » (plutôt que l'imitation directe): la compréhension du comportement d'autrui implique un type de simulation motrice de la même action effectuée par le moi. Dans cette optique, l'imitation comportementale découlerait d'un système semblable de compréhension de l'action. Que cette théorie constitue ou pas l'explication correcte de l'imitation (Jacob & Jeannerod, 2005) et que l'imitation vocale, chez l'être humain, implique ou pas des neurones miroirs, il faut reconnaître que : 1) il doit bien exister un système dans le cerveau humain dont la fonction est de traduire les sons perçus en cibles motrices pour la phonation et l'articulation; 2) que ce système fonctionne à une vitesse incroyable; 3) qu'il constitue, parmi les primates, un trait spécifique de l'espèce des êtres humains. Dès lors, le précurseur de la musique et du langage tel que nous le proposons, basé sur l'hétérophonie contagieuse, devrait être considéré comme un système stéréotypé présentant un répertoire sonore restreint génétiquement et une capacité minimale de mise en correspondance directe par rapport aux hauteurs. Les êtres humains ont ensuite développé une capacité d'imitation phonatoire et articulatoire qui a permis non seulement la création d'un répertoire sonore important, mais également une phénoménale aptitude à l'imitation vocale tout au long de leur vie. Selon moi, la partie inférieure de l'aire de Broca (l'opercule frontal) est une composante primordiale du système impliqué dans le mariage des hauteurs pendant l'imitation vocale, en d'autres termes un système qui reçoit de l'information au sujet des hauteurs perçues et génère des cibles phonatoires en fonction de ces hauteurs.

Enfin, lorsqu'on parle d'imitation, il est utile de garder à l'esprit la distinction entre séquentialité et simultanéité. Au sens conventionnel du terme, l'imitation concerne l'imitation de l'action d'une autre personne *après qu'elle a réalisé une activité*; en d'autres termes, l'imitation concerne la séquentialité. Toutefois, lorsqu'il s'agit d'activités dont les

schémas moteurs figurent, pour tous les participants, dans la mémoire à long terme (comme, par exemple, le fait de chanter « Joyeux Anniversaire »), l'imitation peut simplement impliquer la convergence de tous les participants vers le schéma particulier initié par un individu unique. Donc, bien qu'il puisse s'agir d'une situation de simultanéité, l'imitation peut se révéler importante dans les premiers stades où on établit le schéma (comme par exemple la tonalité à utiliser). Idem pour les slogans qu'on scande ou les chants qu'on entonne dans certaines manifestations sportives. Il est possible que l'imitation puisse simplement s'appliquer aux processus de déclenchement et de convergence.

LA COORDINATION TEMPORELLE : INTÉGRATION ET ALTERNANCE

S'il est évident que l'imitation vocale au cours de l'enfance est importante pour l'imprégnation acoustique, que ce soit pour la musique ou pour le langage, cette même imitation vocale à l'âge adulte devrait revêtir plus d'importance en matière musicale qu'en matière de langage. C'est sans doute dû au fait que la musique comprend souvent une *intégration* de parties (comme lors d'un chant monophonique), tandis que le langage implique généralement l'*alternance* des parties. Une des hypothèses principales présentées dans cet article est que les deux types de configurations de communication se sont développés à partir d'un précurseur commun de vocalisation hétérophonique, un précurseur dans lequel l'arrangement des parties est imprécis (cf. Figure 2). Les arrangements distincts des discours musical et langagier diffèrent principalement au niveau de leurs mécanismes de coordination temporelle. Exprimé plus simplement, un système qui a recours à l'intégration requiert un timing relatif au *déclenchement* des événements, tandis qu'un système qui a recours à l'alternance requiert un timing des *clôtures* d'événements. Nous allons tour à tour examiner ces deux systèmes.

Intégration verticale

Le problème principal qu'on rencontre lorsqu'on passe d'un stade d'hétérophonie contagieuse au stade très synchronisé de la monophonie, de l'homophonie ou de la polyphonie, est celui de l'alignement temporel des pulsations afin de créer un output synchronisé. Comme mentionné précédemment, le chevauchement des voix dans les vocalisations hétérophoniques, comme les hurlements des loups, présente bien des caractéristiques de mélange, mais pas de synchronisation des parties comme on en trouve dans les chœurs formés par des humains. Toutefois, tout comme dans le cas du chant monophonique de type « Joyeux Anniversaire », où il faut que quelqu'un donne le ton pour que les autres participants puissent s'accorder sur lui, le système a également besoin d'un régulateur qui génère à la fois un rythme et un tempo. On peut, comme lorsque nous avons parlé de correspondance de hauteur, penser que ce matching temporel s'obtient au moyen de la simultanéité ou de la séquentialité. On ne pourrait, par exemple, pas imaginer l'apprentissage de nouveaux rythmes de percussions ou de danses sans qu'il y ait capacité générale d'imitation motrice. On est là dans le domaine de la séquentialité. Par contre, lorsqu'il s'agit d'emboîter le pas d'un régulateur existant (comme, par exemple, la synchronisation de ses mouvements avec un rythme de percussion), on se trouve dans le domaine de la simultanéité. La question de savoir si se joindre à un rythme nécessite une capacité d'imitation chez l'humain reste ouverte. La seule certitude, c'est que cette capacité se trouve rarement dans la nature. Parmi les mammifères, elle semble bien n'être présente que chez l'homme, et il s'agit d'une capacité qui a vraisemblablement évolué parallèlement pour la danse et la musique.

Mes collègues et moi-même avons mené récemment une enquête consacrée aux régions du cerveau impliquées dans la synchronisation métrique, dans le contexte de mouvements de danse (Brown, Martinez & Parsons, 2006). Dans cette étude, les sujets devaient bouger leurs pieds, selon un modèle défini,

sur une surface plate et inclinée ; pendant ce temps, leur cerveau faisait l'objet d'une tomographie par émission de positrons. Nous mentionnerons ici deux conditions expérimentales : une dans laquelle les sujets faisaient bouger leurs pieds au rythme de la musique, et une autre dans laquelle ils bougeaient leurs pieds sans musique, mais avec *grosso modo* le même tempo. Si les deux tâches faisaient bien intervenir des parties presque identiques du système moteur de contrôle des extrémités inférieures, la première d'entre elles faisait apparaître une forte augmentation de l'activité du vermis médian du cervelet ; dans la seconde tâche, l'activation de cette même zone était moins importante. Cette grande augmentation de l'activité du cervelet n'était pas due, et c'est important, à la présence de musique dans la condition de synchronisation : en effet, l'ensemble de l'activité cérébrale due à l'écoute de la musique avait été soustraite après mesure par un scan séparé ; l'intensité de l'activation n'en était pas diminuée. La différence relevée était donc un reflet relativement pur de la synchronisation audiomotrice elle-même. On ne sait pas si le vermis médian est impliqué dans la synchronisation vocale au même titre qu'il l'est dans la synchronisation corporelle. Dans notre étude (déjà mentionnée) consacrée à la vocalisation (Brown *et al.*, 2004), on relevait une activité de bas niveau dans cette région du cerveau lorsque les sujets devaient entrer en synchronie avec une autre partie musicale (soit une forme de synchronisation vocale), mais pas lorsqu'ils devaient reproduire des mélodies (soit une forme d'alternance et d'imitation vocale). Au niveau conceptuel, on devrait s'attendre à ce qu'un régulateur neural fonctionne comme une entité supérieure agissant sur divers types de signaux sensoriels et contrôlant de multiples effecteurs moteurs : en effet, les humains peuvent synchroniser leurs mouvements en fonction de sons, de flashes de lumière, de pressions sur la main, *etc.* et peuvent engager les mouvements de presque toutes les parties de leur corps par rapport à ces signaux. Il est au minimum possible que le vermis médian soit, chez l'homme, une zone cérébrale qui intervient dans la synchronisation métrique.

Il vaut la peine de mentionner qu'on a observé, chez certains insectes et espèces de grenouilles, des appels qui font penser à de la synchronisation (Greenfield, 1994 ; Pollack, 2000) ; dès lors, on peut penser que l'intégration ne nécessite qu'une configuration neurale minimale. Si on part de ce principe, on peut s'étonner du fait que les appels intégrés n'ont pas connu un développement plus important dans le monde animal, et qu'on ne les trouve même pas chez des espèces capables de vocalisations complexes (comme les oiseaux chanteurs ou les baleines à bosse). Ceci suggère une dissociation entre, d'une part, l'imitation phonatoire et, d'autre part, la synchronisation phonatoire : 1) l'imitation phonatoire existe chez les oiseaux chanteurs et les baleines à bosse, en l'absence de synchronisation phonatoire ; 2) la synchronisation phonatoire existe chez les insectes et les grenouilles en l'absence d'imitation phonatoire ; et 3) les humains présentent, dans leurs vocalisations, un cas unique de combinaison d'imitation phonatoire et de synchronisation métrique. Il est possible que l'imitation de la hauteur et la mise en correspondance par rapport à cette hauteur (processus qui impliquent, respectivement, de la séquentialité et de la simultanité) requièrent la mise en œuvre de processus cognitifs différents ; il est également possible que les deux caractéristiques reposent conjointement sur la capacité d'imitation de l'être humain. À nouveau, il semble évident que les modèles temporels d'imitation séquentielle requièrent, au départ, la présence d'une capacité d'imitation. D'autres recherches seront nécessaires afin de déterminer si c'est également le cas pour la synchronisation par rapport à une pulsation en temps réel.

Si on veut résumer les idées relatives à la musique présentées dans les deux parties précédentes de cet article, on voit que le glissement d'un système pré-curseur de vocalisation hétérophonique au système moderne d'intégration musicale requiert le développement d'au moins deux capacités clés : 1) la capacité de mise en correspondance avec les hauteurs, avec médiation partielle par l'opercule frontal et 2) la synchronisation (entrainement) avec médiation

partielle du vermis médian. Il est possible que l'opercule frontal joue également un rôle dans le traitement rythmique en plus de celui qu'il joue dans l'imitation des hauteurs, comme l'ont montré les études (citées plus haut) qui pointent son activation lors de tâches de discrimination en rapport avec le rythme et la durée. De plus, le cervelet, qui joue un rôle dans la synchronisation, peut également en jouer un dans l'imitation ; en effet, Penhune, Zatorre et Evans (1997) ont montré que le cervelet (y compris le vermis) était activé lorsque des sujets devaient imiter des schémas rythmiques acoustiques en tapant du doigt.

Au-delà de cette discussion relative aux capacités cognitives fondamentales, je pense que la musique a adopté une approche minimaliste du processus de génération de phrase. Par rapport au langage, la musique a eu tendance à devenir un système simple, qui se base surtout sur la répétition. Lomax (1968) a fait depuis longtemps remarquer que le chant constituait de loin le plus redondant des systèmes de vocalisation chez l'être humain (*cf.* également Richman, 2000). La répétition est certainement une des caractéristiques les plus évidentes de la forme musicale, quelle que soit la culture d'où elle est issue. Un bon exemple de ce processus est constitué par les systèmes basés sur l'*ostinato*, comme par exemple la polyphonie africaine (Arom, 1991). La combinaison de la synchronisation métrique et de la redondance de phrase fait alors de la musique un système *cyclique* et, dans ce système, la répétition ne se trouve pas seulement dans le recours à des intervalles isochrones mais également dans la répétition des phrases. En matière de musique, ces deux processus sont primordiaux dans l'établissement de l'intégration. D'un autre côté, ils limitent probablement le contenu informatif général de la musique, dans la mesure où le système est plutôt orienté «répétition» que «nouveau». D'un point de vue social, l'intégration verticale de parties constitue un outil idéal de promotion de la communication de groupe, dans la mesure où l'objectif de ce type de processus est la création d'un output acoustique commun et coordonné. Ce moyen relati-

vement égalitaire de production musicale permet à tous les participants d'avoir des rôles équivalents, ce qui entretient la symétrie de ces rôles. Et c'est optimal dans une optique de cohésion, de coopération et de catharsis du groupe (Brown, 2000b).

Alternance horizontale

Si l'intégration verticale est l'outil essentiel de la musique en groupe, l'alternance constitue le mode dominant de discours langagier, mode qui se retrouve également dans de nombreuses formes de musique chorale (principalement les formes antiphonique et responsoriale). À ma connaissance, la forme principale de langage intégré verticalement est constituée par les slogans que l'on scandie lors des meetings politiques. Mais comment arriver, à partir du système à forte simultanéité présenté par le précurseur hétérophonique, à un système de véritable alternance, tel que celui qui sous-tend la conversation chez l'être humain ? Le problème de base posé par le précurseur hétérophonique est celui d'un sentiment de confusion sémantique dû au chevauchement des parties, comme lorsque de nombreuses personnes parlent en même temps. L'objectif de base d'un système d'alternance sera donc d'opérer une *ségrégation des parties*, de manière à ce que le chevauchement soit interdit et que l'alternance devienne la règle de la communication. L'alternance peut se produire de diverses manières, mais le moyen idéal serait de rendre les parties individuelles contiguës en mettant en œuvre un processus qu'on pourrait appeler «chaînage» (*stringing*). Par le chaînage, non seulement les diverses parties sont à la fois séquentielles et séparées, mais une partie commence à l'endroit où une autre s'arrête, de telle sorte qu'il existe un flux continu d'interaction : les intervalles et les chevauchements y sont minimes ; en d'autres termes, la coordination se fait par rapport aux clôtures d'événements. Notons que, si on le compare au système d'intégration (dont la nature est cyclique et redondante), un système d'alternance est fortement linéaire et génératif. Ce type de système convient

particulièrement à l'échange d'informations ; on ne pourrait le concevoir dans un système cyclique basé sur la répétition. En tant que tel, il s'appuie fortement sur l'*asymétrie* des rôles, d'où la différenciation des rôles comme celle qu'on trouve dans les formes de musiques responsoriales, dans la distinction entre soliste et chœur. À nouveau, cette diversité des parties est idéale pour l'échange d'informations. Les dyades (qui constituent la configuration sociale typique du langage chez les humains) tendent à favoriser l'alternance. Malheureusement, il n'y a pas, à ma connaissance, d'étude par neuro-imagerie qui ait explicitement examiné, chez l'homme, la configuration temporelle sous-tendant l'alternance, que ce soit par rapport au langage ou à la musique. La plupart des gens diraient vraisemblablement que l'alternance est plutôt régulée par des considérations de nature *sémantique* qu'uniquement par des signaux acoustiques. Les chercheurs qui consacrent leurs travaux à l'analyse du discours ont tendance à se focaliser sur les signaux sémantiques qui mènent à un compromis par rapport aux rôles de locuteurs plutôt que sur les signaux acoustiques uniquement (Yule, 1996). Il y a là sans nul doute matière à d'autres études. Il est intéressant de constater que de nombreuses formes de duos chez les animaux impliquent des processus basiques d'alternance (par exemple, ceux des pies australiennes, cf. Brown & Farabaugh, 1991). Selon Geissmann (2002), qui suit en cela Wickler et Seibt (1982), il est possible que l'alternance observée dans les duos de gibbons soit issue d'un processus appelé « song splitting » – littéralement « séparation du chant » – par lequel un chant commun du mâle et de la femelle se divise ensuite entre les deux partenaires, et qu'il s'agisse là d'un mécanisme raisonnable par lequel l'alternance a pu se développer chez l'être humain, un mécanisme tout à fait compatible avec mon modèle de précurseur hétérophonique.

Dans la Figure 3, on trouvera une présentation de trois voies différentes qui peuvent avoir été empruntées à partir de ce système précurseur basé sur l'hétérophonie contagieuse, pour en arriver aux formes contemporaines de la communication vocale

chez l'être humain ; le schéma montre bien que le système d'intégration (système 1) est un système fait de correspondance de hauteurs, de synchronisation métrique, de redondance, de rassemblement, de cohésion sociale et de symétrie des rôles ; il s'oppose au système d'alternance (système 3) qui, lui, est un système de séparation, de chaînage, de générativité, de dyades, d'échange d'informations et d'asymétrie des rôles. Lorsque j'avance l'hypothèse selon laquelle le discours musical et le discours linguistique proviennent d'un système hétérophonique conjoint, je dis en fait qu'intégration et alternance sont des variations sur un même thème. Et, en fait, ces deux formes de discours cohabitent, à travers le monde, dans de nombreux contextes musico-linguistiques, comme le montre un autre voie importante, qui serpente entre les deux extrêmes formés par l'intégration et l'alternance : c'est la voie de l'*alternance avec intégration* (système 2). Cette forme de voie est celle qu'emprunte la configuration soliste/chœur qu'on trouve dans diverses formes de musique chorale partout dans le monde. La Figure 3 montre bien que, dans ce type de configuration du discours, le chœur peut être très intégré (comme dans le cas de l'intégration pure) ou très hétérophonique (comme dans le cas du précurseur hétérophonique). De plus, l'interaction du soliste et du chœur peut être quasiment nulle (comme dans le cas de l'alternance pure) ou présenter un chevauchement important (comme dans le cas du précurseur hétérophonique). Il est important de noter que lorsque le chœur est très intégré et que les interventions du soliste sont fortement séparées de celles du chœur, alors la configuration soliste/chœur devient celle d'une pseudo-dyade, en d'autres termes une approximation groupale de deux individus en train de converser. D'un point de vue social, la voie de l'alternance avec intégration se trouve, en termes de « distinction groupe/dyade » et « spectre de symétrie/asymétrie des rôles », quelque part entre celle de l'intégration pure et celle de l'alternance pure (cf. Figure 3). Par exemple, à un des extrêmes, nous avons des chœurs très intégrés et qui récitent exactement le même refrain en réponse à chacun des appels du soliste, eux-mêmes variables pour

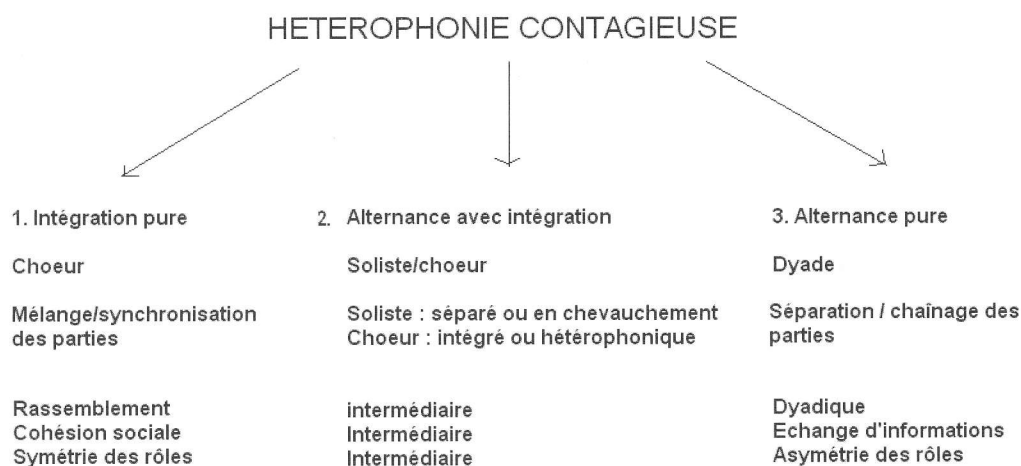


Figure 3.

Comparaison de trois formes de configuration du discours dans la communication entre humains : intégration, alternance, et alternance avec intégration.

chaque couplet. C'est bien le soliste qui fournit une information nouvelle à chaque couplet, pas le chœur (intégré).

CONCLUSIONS

Dans cet article, j'ai présenté quelques idées basées sur l'importance qu'a revêtu le « son » (sounding) dans les origines de la musique. Parler d'harmonie, c'est parler des processus interpersonnels de l'exécution musicale, par lesquels des individus mélangent, volontairement et de manière coopérative, des lignes musicales dans l'espace (hauteurs) et dans le temps. J'y ai également présenté un précurseur vocal possible du chant et du langage chez l'être humain, basé sur l'hétérophonie contagieuse et ai suggéré qu'un système de matching audiovocal situé dans l'aire de Broca fournit les fondements neuraux des processus de mise en correspondance des hauteurs. J'ai également suggéré que le vermis médian joue un rôle dans la synchronisation métrique pour l'intégration des diverses parties. S'il est vrai qu'on peut trouver les

rudiments des systèmes tonals dans les vocalisations hétérophoniques contagieuses comme les hurlements des loups, ce qui me paraît plus important, c'est la configuration groupale de ce type de vocalisations plutôt que leurs propriétés musicales. Je crois qu'un chœur de singes hurleurs, aussi peu uni qu'il soit dans les hauteurs, et avec une bande passante aussi large, nous apprend plus au sujet des origines de la musique que le chant d'oiseau le plus mélodieux. Enfin, j'ai discuté de la possible émergence du discours musical et langagier au travers de divers types de processus, qui se concentrent soit sur l'intégration (principalement la musique), l'alternance (principalement le langage) ou l'alternance avec intégration (principalement la musique à base de langage). De manière générale, je considère l'harmonie comme une conséquence acoustique directe du fait que nous faisons partie d'une espèce coopérative et vivant en groupe. Et quel type de processus de communication pourrait être plus important pour la collectivité culturelle de l'être humain que le mélange coordonné de voix qui soutient le chœur des humains ?

Traduction Stéphane Renard