

*Colwyn Trevarthen, Maya Gratier*

VOIX ET MUSICALITÉ :  
NATURE, ÉMOTION, RELATIONS ET CULTURE

LA VOIX, INSTRUMENT DE SOCIABILITÉ

La voix a cette caractéristique particulière d'être tournée à la fois vers l'intérieur et vers l'extérieur ; d'être à la fois physique et psychique. La voix passe de l'un à l'autre, elle est véhicule, mais aussi vécu véhiculé, forme et substance. Elle exprime les variations infinies d'états psychiques, variations chargées de sens expressif. Les expressions vocales sont organisées en paroles verbales et sont affectées d'une musicalité spontanée aussi significative que les mots. Le bébé qui découvre la potentialité expressive de sa propre voix apprend en même temps à jouer avec le temps du partage dans lequel elle s'inscrit et à maîtriser les limites de l'espace entre le « je » et le « tu ».

L'idée que le bébé bouge et répond de manière rythmique et musicale à la riche mélodie de la voix maternelle, qui partage les caractéristiques de l'art musical, gagne du terrain parmi les psychologues de la petite enfance (Trevarthen, 1999). Si nous acceptons cela, alors la voix doit être le premier instrument musical au cours du développement ; elle a d'ailleurs certainement fait partie de la première performance musicale créée par l'espèce humaine (Cross, 1999 ; Morley, 2002). L'origine de la musicalité humaine a été liée au « chant » des animaux, mais beaucoup de caractéristiques lui sont uniques. Les jeux expressifs, calmes et beaux, qui engagent l'affection d'un proche pour un bébé portent la curiosité et l'appétit affectif de l'enfant pour l'apprentissage du monde de ses ancêtres.

LA VOIX DE LA VITALITÉ INTIME, SOURCE D'AFFECTION  
DANS UN CERVEAU MUSICAL

Les talents musicaux du bébé ont récemment été démontrés scientifiquement (Fassbender, 1996 ; Papousek, 1996 ; Trehub, 1990). Doté d'une capacité d'écoute inattendue, le nourrisson montre un intérêt particulier pour la voix humaine. Les Papousek ont décrit le « parentage intuitif » qui s'appuie sur des phénomènes musicaux comme la modulation affective par la mélodie et par le rythme (Papousek et Papousek, 1981). Une étude chronologique de leur fille documente son plaisir à écouter les chansons pour bébés au cours de sa première année ainsi que sa pratique privée de formes musicales acquises. Daniel Stern (1999, 2000) décrit l'accordage affectif du parent lorsqu'il reprend l'expression vocale du bébé en reflétant son rythme, ses accents et intonations par une autre modalité expressive.

Un nouveau-né, même s'il ne comprend aucun mot, semble sentir les expressions motivées variées de la voix de l'adulte et il sympathise avec elles (Fernald, 1992). En effet, nous savons à présent qu'un fœtus peut écouter le son de la voix maternelle transmis à travers son corps et qu'il apprend à connaître sa signature vocale particulière dans les semaines qui précèdent sa naissance. Un jeune bébé est capable de bouger et de se synchroniser au timing de la parole ludique et affectueuse, et peut exprimer une gamme d'états d'animation, allant de la surprise et l'intérêt, à la détresse et à la sympathie (Jaffe et coll., 2001). Et tout cela, sans aucun entraînement pour les conventions et usages que nous définissons comme le langage parlé.

Nous pensons que la « musicalité communicative » que les bébés révèlent, d'une part à travers leurs sons expressifs et d'autre part en raison de leur sensibilité pour les sons d'autres personnes, constitue le fondement essentiel pour la maîtrise progressive du langage parlé (Malloch, 1999 ; Papousek et Papousek, 1981 ; Trevarthen et Malloch, 2002). La musicalité innée du cerveau, et son sens du temps, de l'expression mélodique et de la narration motivent le corps qui chante et danse. La bande de fréquence du langage est réglée par des muscles spéciaux de l'oreille moyenne, capables d'amortir le bruit qui détourne l'attention de la parole significative. À travers la voix affectueuse du parent, le jeune bébé apprend à partager ses émotions et à identifier ses buts et ses idées.

LA VOIX DU BIEN-ÊTRE ET LE PARTAGE NARRATIF

De toutes les expressions du corps, celles produites par les vibrations des cordes vocales sont le plus intimement liées à la régulation interne de

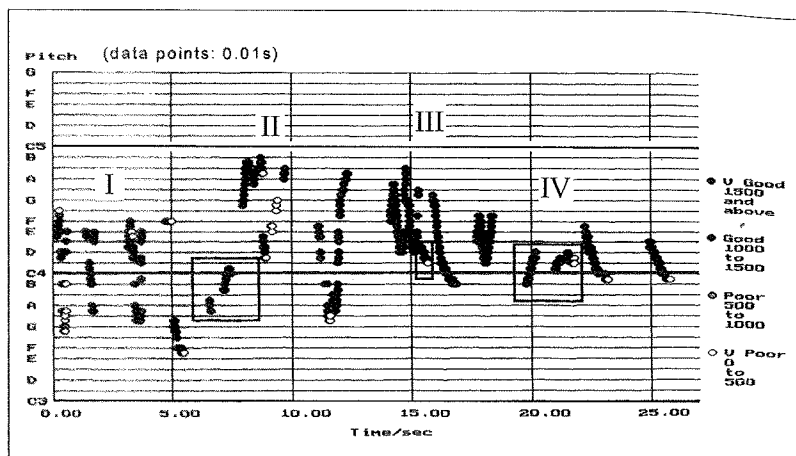


Figure 1. Le cycle narratif dans la vocalisation maternelle. Les gestes vocaux tracés par les changements de la hauteur de la voix d'une mère écossaise parlant affectueusement avec sa fille de six semaines, Laura. La parole de la mère ressemble à un chant présentant une introduction calme (I), une animation croissante après la réponse du bébé (premier rectangle) (II), un apogée d'excitation où des vocalisations sont échangées avec rapidité (III), et finalement un dénouement calme et lent organisé autour d'une fréquence centrale (de 260 Hz) où la mère imite deux fois la vocalisation du bébé (IV). Juste avant l'apogée, la hauteur de la voix de la mère se situe à une octave au-dessus du reste du dialogue (Malloch, 1999).

nos états de vitalité et de bien-être. Les remarquables pouvoirs de signalement social de l'appareil vocal du mammifère ont évolué à partir de structures qui, chez des animaux comme les poissons et les reptiles, servent à réguler les états internes et l'économie d'énergie, en contrôlant l'ingestion d'oxygène, d'eau et de nourriture (Panksepp et Bernatzky, 2002 ; Porges, 2003 ; Scherer, 1986). Le pouvoir et la qualité du contrôle de nos actions intentionnelles sont perçus dans la voix qui est capable de signaler les mouvements instantanés d'états d'animation et de santé. Mais ce n'est pas tout. La vocalisation humaine se construit autour d'un besoin de raconter une histoire, et cela nous conduit vers l'univers de la signification.

La voix peut raconter des histoires d'intentionnalité et refléter les expériences passées en reproduisant une « narration » d'intention. Comme un

oiseau ou un singe gibbon qui chantent sur leur terrain (Marler et coll., 1992 ; Merker, 2000), l'être humain tisse un message de communication, un pattern rituel d'activité motrice à laquelle d'autres peuvent répondre. C'est ainsi que les membres d'un même groupe partagent leur expérience d'un monde commun et de leur activité intentionnelle. Et bien plus que pour les autres espèces sociales, l'histoire humaine est élaborée sous forme de narration liant espoirs remémorés et souvenirs, goûts et dégoûts, craintes et fantaisies, une mise en scène de métaphores mimétiques pour des personnalités imaginées s'attribuant des rôles dans des espaces et des temps imaginés. Ainsi est créé le fondement syntaxique des formes linguistiques : une grammaire d'expressivité significative qui sous-tend nos mythes et légendes, tous nos arts temporels, la parole et l'écriture.

Une personne qui chante, récite une poésie ou parle afin de transmettre des informations propose un partage qui est avant tout une aventure. C'est cela qui fait que le chant, la parole et toute forme expressive humaine sont des phénomènes fondamentalement musicaux, et potentiellement culturels. La musique est l'un des arts de l'intimité qui raconte notre histoire, et la voix est la plus précoce et la plus variée de ses véhicules (Dissanayake, 2000 ; Stern, 1999).

La recherche fondée sur l'analyse du jeu vocal interactif entre parents et bébés, recherche qui retrace la façon dont le bébé gagne confiance et habileté dans le domaine linguistique, nous a amenés à définir une théorie de la « musicalité communicative » (Malloch, 1999 ; Trevarthen, 1999). Cette méthode descriptive fine de l'histoire naturelle humaine montre la façon dont le timing et la modulation énergétique du mouvement expressif sont essentiels pour la communication de narrations actives, pensées et ressenties qui « racontent » quelque chose qui ne peut exister en dehors de l'esprit humain. La musicalité, avant la naissance et au sein du langage, est le véhicule d'un « sens commun », qui est un savoir partagé facilitant l'apprentissage culturel (Dissanayake, 2000).

La musicalité communicative est définie selon trois dimensions : la *pulsation*, la *qualité* et la *narration* (Malloch, 1999).

La *pulsation* est la succession régulière d'étapes comportementales discrètes dans le temps, représentant un processus de « création du futur » par lequel un sujet peut anticiper ce qui arrive et quand.

La *qualité* est constituée par les contours vocaux et les gestes corporels expressifs, donnant forme au temps en mouvement.

La *narration* de l'expérience individuelle et de la camaraderie est construite à partir de la séquence des unités de la pulsation et de la qualité que l'on trouve dans les gestes créés en commun – elle décrit la façon dont ils sont reliés pour former des chaînes d'expressivité.

## LA VOIX DE L'AVENTURE DANS LES CHANSONS POUR BÉBÉS

Dans le chant, la voix exprime l'esprit de l'action et de la découverte humaine. L'attention du bébé est captivée par sa mère lorsqu'elle chante pour lui et cette entente constitue le noyau dur de la musicalité élaborée de musiciens habiles, capables de transférer leurs impulsions créatives à leurs instruments musicaux (Dissanayake, 2000 ; Trehub, 1990 ; Trevarthen, 1999). La musique, toute musique, donne voix au mouvement coordonné des parties du corps et à la manière dont l'esprit poursuit ses trajectoires par le mouvement (Clynes, 1980).

La participation intéressée d'un bébé de 6 mois au chant de sa mère, sa réponse coordonnée et précise aux moments clés de la narration musicale, et en particulier aux voyelles longues qui riment pour marquer la fin des phrases, montrent que le plaisir de la mère met en relief le besoin du bébé pour ce type d'action partagée (Trevarthen et coll., 1999 ; Trevarthen et Malloch, 2002). Les chansons des parents et des puéricultrices varient dans leur degré d'animation et d'énergie ; elles peuvent servir à amuser

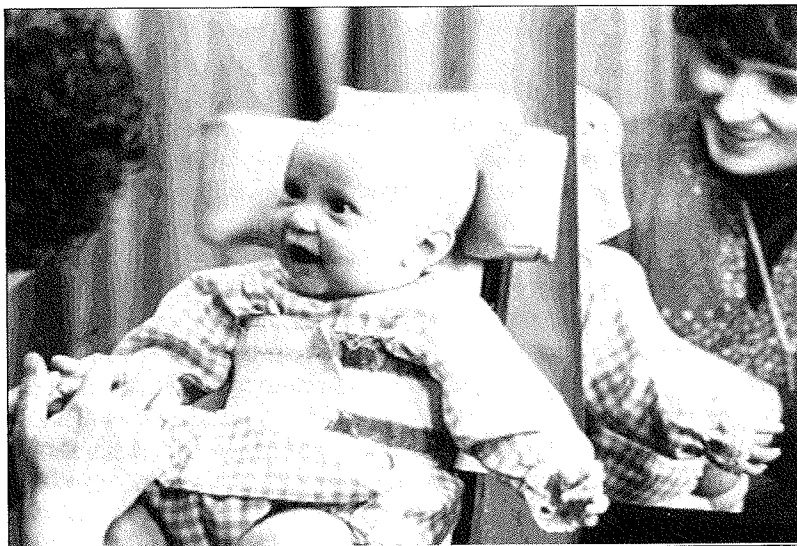


Figure 2. Leanne, 5 mois, à Édimbourg. Elle suit avec enthousiasme la récitation d'un chant d'action par sa mère (*Round and Round the Garden*). La mère décrit des cercles sur la main de Leanne, fait monter ses doigts sur son bras et la chatouille soudainement sous le bras. Leanne connaît bien ce chant et vocalise souvent sur la dernière voyelle longue de sa mère.

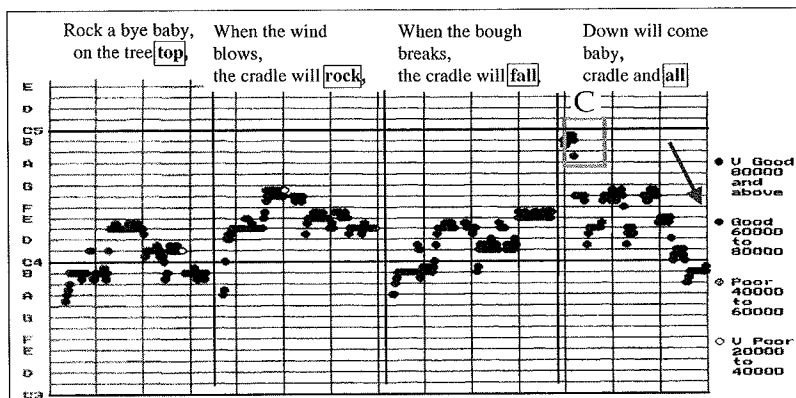


Figure 3. Les changements de la hauteur de voix d'une femme qui chante la berceuse *Rock a Bye Baby*. Ce verset comprend quatre vers, qui durent chacun 28 secondes. Les voyelles longues riment à la fin des lignes 1 et 2, et 3 et 4, et ces sons attirent l'anticipation du bébé et sont souvent imités. L'apogée de la chanson (C) a lieu au début de la ligne 4, avec le mot *down*. Il est suivi par une baisse précipitée de la hauteur du son.

ou à exciter un bébé, ou bien à le calmer et à le conforter lorsqu'il est fatigué ou mal à l'aise. L'invention de jeux taquins entre parents et bébés, qui implique un élément de risque et de surprise calculé, montre que la modulation de la vocalisation peut constituer un outil social régulateur des intentions et des joies de la découverte synchronisée.

Le bébé apprend les chansons et les actions qui les accompagnent, et s'approprie les jeux les plus passionnants pour en faire des performances publiques mises en scène avec fierté pour les autres (Trevvarthen, 2002). Cette fierté pour la performance de jeux musicaux familiaux constitue une preuve solide que nos motivations pour la musique comme instrument d'un savoir communautaire sont innées. L'étude comparative de chansons dans différentes langues soutient l'hypothèse de patterns musicaux universels qui sous-tendent les formes culturelles variées des chansons.

#### LA VOIX DE LA TRISTESSE ET DE LA SOLITUDE

L'analyse acoustique de « conversations » et de jeux chantés entre mères et bébés démontre les effets de la maladie psychique de la mère sur sa voix et sur la vitalité et l'expressivité du bébé (Robb, 1999). Nous avons trouvé qu'il est possible de cerner et de comprendre les mécanismes de la

remarquable efficacité de la thérapie musicale en analysant comment l'engagement musical improvisé peut tirer un individu du retrait psychique ou de l'antagonisme social (Trevarthen, 2001).

La musique que les mères partagent avec leurs bébés porte des liens particulièrement valorisés par la mère : des liens avec sa famille, sa communauté et ses expériences passées. Les mères chantent leur place au sein d'un lieu aimé, et elles chantent aussi pour se reconforter et assouvir leurs peurs et leur solitude, prêtant ainsi au bébé le rôle de confident ou de gardien de leurs secrets.

Les émotions de la protoconversation ont été testées en prêtant attention à ce qui se passe lorsque la réponse humaine à l'intérêt du bébé fait défaut. Si une mère fige l'expressivité sympathique et spontanée de son visage pendant une minute alors qu'elle joue avec son bébé de 2 mois, celui-ci manifeste des émotions claires de retrait et de détresse (Tronick et coll., 1978 ; Murray et Trevarthen, 1985). Un pattern semblable d'anxiété et de tristesse est apparent lorsque la mère présente un style de communication déprimé. La dépression postnatale entrave la communication et, si elle perdure, elle s'accompagne souvent chez le bébé de troubles du développement cognitif (Murray et Cooper, 1997 ; Tronick et Weinberg, 1997). Un adulte malheureux et peu communicatif ne peut être un bon compagnon ni un bon enseignant.

Dans une autre étude, grâce à l'utilisation d'un appareil à double téléviseurs, le bébé et la mère sont placés dans deux pièces séparées et interagissent en face à face à travers un système vidéo et sonore commun (Murray et Trevarthen, 1985 ; Nadel et coll., 1999). Cet appareil permet de comparer minutieusement les réactions du bébé lorsque l'interaction a lieu en temps réel et lorsqu'elle est artificiellement décalée d'une demi-seconde. Les expressions de la mère décalées dans le temps engendrent des comportements de détresse chez le bébé, soutenant l'idée qu'un bébé de 2 mois est extrêmement sensible à la contingence des expressions sociales. Et une mère à laquelle on présente un enregistrement de son bébé qui est légèrement décalé dans le temps décrit un sentiment de malaise et de culpabilité lié à un manque de contact. La communication vive doit donc impliquer un engagement mutuel de sentiments et d'impulsions expressives.

#### LA VOIX DE L'APPARTENANCE DANS LES DIALOGUES ENTRE MÈRES ET BÉBÉS

La musicalité expressive de la protoconversation représente un véhicule universel pour le partage d'expériences culturelles. Le partage d'un savoir commun est sous-tendu par la motivation innée du bébé à par-

tager son vécu avec ses proches à travers la coordination dynamique et l'anticipation de gestes expressifs. À travers cette coordination des expressions vocales, visuelles, gestuelles et posturales, la mère et le bébé partagent leurs vécus de manière directe et immédiate au sein d'un espace psychique intersubjectif. L'intersubjectivité implique une conscience sympathique et une coordination des motivations de chacun. Le bébé naît avec un esprit ouvert, tourné vers l'autre, et tout être humain se situe d'emblée dans des espaces qui sont culturels.

La méthodologie nouvelle fondée sur l'analyse acoustique détaillée d'expressions vocales spontanées nous permet de cerner la dynamique expressive de l'interaction vocale précoce. L'analyse de trente-six interactions naturelles mère-bébé dans trois contextes culturels (France, Inde, États-Unis) montre que la mère et le bébé accordent d'instant en instant leurs expressions vocales de manière quasi musicale ou avec une « musicalité » suivant une structure temporelle cohérente et universelle. Il est évident également que les interactions vocales spontanées portent déjà les empreintes d'appartenance culturelle. Nous avons pu démontrer ce lien entre les espaces culturels de la mère, ses représentations et comportements, ses habitudes, ses appartenances, et l'espace intersubjectif qui

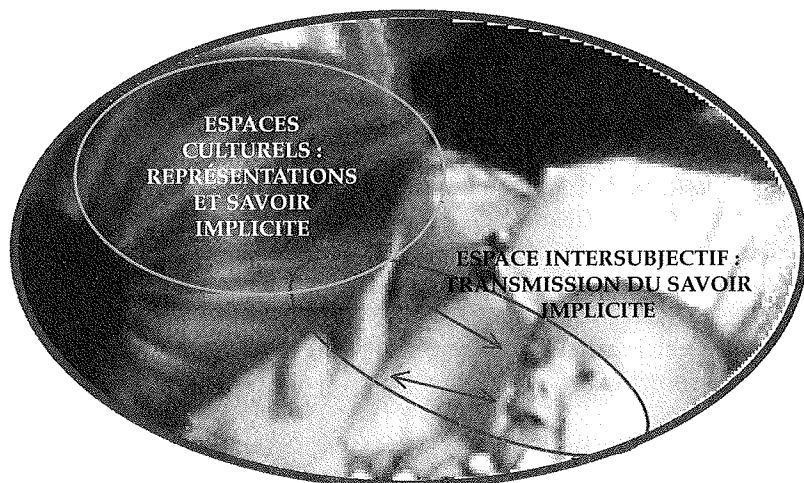
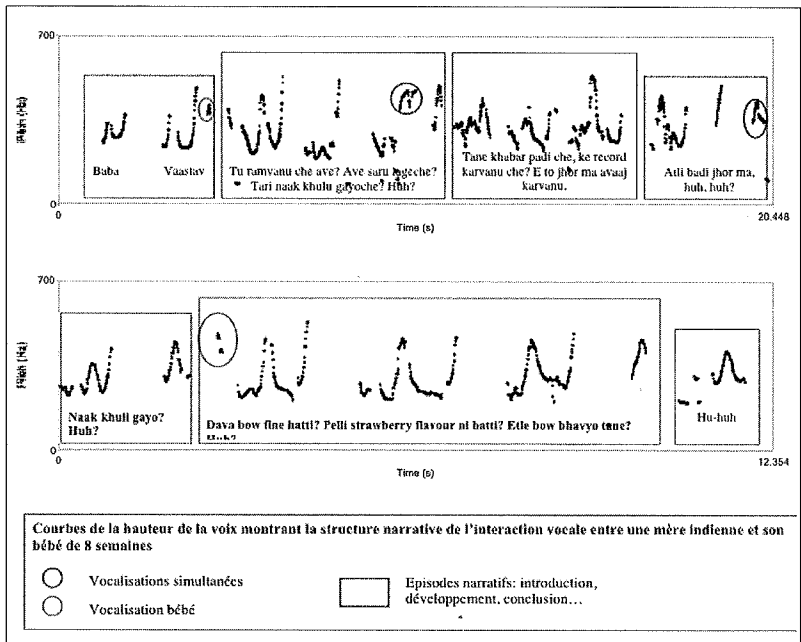


Figure 4. Comment les espaces de l'intersubjectivité et de la culture se rencontrent au cours de l'interaction mère-bébé.

s'établit entre elle et son bébé à travers la dynamique interactive. L'expérience culturelle à la fois façonne et est elle-même coconstruite au sein de l'espace intersubjectif mis en place et maintenu par la coordination sensible et subtile des expressions de la mère et du bébé.

Les bébés que nous avons étudiés avaient tous entre 8 et 16 semaines, étaient les premiers-nés, nés à terme et en bonne santé. Des enregistrements audio furent effectués au domicile, lors d'interactions spontanées en face à face entre la mère et le bébé. Soixante secondes d'interaction vocale sont sélectionnées pour chaque dyade. Des représentations spectrographiques et des courbes de la hauteur des voix ont permis de procéder à une analyse du « timing » et des formes prosodiques des vocalisations des mères et des bébés.

Nous avons trouvé, d'une part, que les dyades dans les trois contextes culturels présentent une organisation temporelle similaire dans la coordination expressive de leurs vocalisations. Toutes les



Figures 5 et 6. Exemples d'analyse acoustique d'une dyade française et d'une dyade indienne montrant des « signatures prosodiques » culturelles et la construction narrative universelle de l'interaction.

interactions sont organisées autour de trois unités temporelles hiérarchiques de durées moyennes semblables : une « pulsation » d'environ une seconde, une « phrase » d'environ trois secondes et un « épisode narratif » d'environ trente secondes. Nous avons également démontré que les dyades se distinguent selon leur origine culturelle du point de vue d'un certain nombre de caractéristiques, comme la durée de vocalisation, la durée des pauses, le degré de verbalisation (Gratier, 2001).

Ces analyses d'interactions vocales spontanées dans les trois pays sus-cités montrent que dès les premières semaines de la vie, les bébés sont déjà imprégnés de leur culture (Hall, 1989) et que leur appartenance culturelle s'exprime dans la microtemporalité des premiers liens affectifs. Ils commencent aussi sans doute à incorporer – c'est-à-dire à inscrire dans le corps – les règles conversationnelles spécifiques et implicites de leur communauté d'appartenance.

En effet, les rythmes et cadences de la conversation varient énormément d'une culture ou d'un contexte à un autre : dans certains contextes il est approprié et attendu que les interlocuteurs parlent en même temps ou se rejoignent dans leurs expressions vocales, verbales et non verbales ; ces pratiques, dans d'autres cultures, enfreignent les règles de la politesse. De la même manière, la pause qu'il est approprié de laisser avant de prendre la parole varie selon les cultures (Kerbrat-Orecchioni, 1994). Mais il faut souligner que ce qui est fondamental pour la protoconversation et qui est probablement ancré dans la structure du cerveau humain, c'est la capacité de générer et de percevoir l'expression organisée dans le temps (Gratier, 1999 ; Trevarthen, 1999).

En s'accordant aux modes intérieurs et aux ressentis fluctuants des autres, ainsi qu'aux façons d'être, de ressentir et de s'exprimer imprégnées par la culture ambiante, le bébé gagne de l'expertise dans le domaine du savoir implicite. Il le fait grâce à une sensibilité multimodale pour les contours temporels de l'expression humaine dont la voix, son énergie et ses dynamiques représentent un véhicule primordial. Nous proposons que le savoir culturel implicite très précoce, inscrit dans le grain de la voix et la microexpressivité du corps, construit un sens profond d'appartenance et soutient l'inscription de l'enfant dans une communauté de proches.

#### LA VOIX DE LA CONNAISSANCE ET DE L'ENSEIGNEMENT

Le timing et la qualité de l'expression dans la communication sont tout aussi importants lorsqu'il s'agit de l'enseignement et de l'apprentis-

sage concernant des enfants plus âgés (Trevarthen, 2004). Erickson a exploré la façon dont le timing et les indices contextuels opèrent dans la salle de classe afin d'aider les participants à prévoir ce qui est susceptible de se produire.

« Ce qui paraît maintenir la cohérence de toute l'écologie de l'interaction au cours de sa performance c'est le timing. La place relative dans le temps des actions diverses des interlocuteurs constitue un élément important de l'ordonnement de l'activité conversante collective dans ses aspects tant réciproques que complémentaires » (Erickson, 1996, p. 34).

Cette interaction est façonnée par des « patterns de cadence » produisant des « points saillants dans le courant comportemental verbal et non verbal ». Enseignants et élèves organisent leur attention de façon collective et contribuent de la sorte à une manière d'écouter et de parler intégrée et cohérente qui est à la fois naturelle, agréable et mobilisatrice. On peut former l'hypothèse que là où l'enseignement est inefficace, le tour de rôle sans rupture inscrit dans un cadre temporel organisé est défaillant.

Une étude effectuée en Écosse a exploré des méthodes d'analyse pour étudier ce que l'on peut appeler le « parler-enseignant », ces formes particulières de langage que les enseignants adressent à de jeunes élèves (Robb et coll., 2003). Des enregistrements vidéo d'interactions entre élèves et enseignants ont été réalisés, suivis d'une analyse fine destinée à dégager les traits essentiels d'une communication qui satisfait et qui atteint son but. Un groupe cible de huit enseignants, sélectionnés par leurs collègues pour leurs compétences communicatives particulièrement bonnes, a été comparé avec un groupe de sept enseignants quelconques, mais ayant le même niveau d'expérience dans l'enseignement. Chaque enseignant choisit six à huit élèves de sa classe primaire pour participer à une discussion de groupe. La tâche imaginaire suivante a été présentée aux élèves. « Vous devez passer une semaine sur une île déserte avec un/une ami(e). Vous devez trouver un trésor caché. Réfléchissez à tout ce dont vous aurez besoin pendant cette semaine. » Cinq minutes des discussions filmées ont été sélectionnées pour une analyse acoustique détaillée. Les catégories analysées incluent le degré de contact, l'humeur, l'initiative verbale et non verbale des enseignants et les réponses aux élèves. Une analyse du discours a permis d'identifier les styles et la fonction des commentaires. Les durées et les fréquences des tours ont été mesurées, ainsi que le timing, la hauteur du son et le ton vocal des enseignants et des élèves.

Les deux groupes d'enseignants ne se distinguent pas du point de vue de la qualité de leurs réponses positives aux élèves, mais les analyses montrent que le groupe d'enseignants cible était plus attentif aux élèves,

plus animé, plus amusant et plus soutenant. Ses réponses présentent davantage d'interventions métacognitives et réflexives. Les spectrogrammes des expressions vocales des participants montrent que le groupe cible présente davantage d'accordage réciproque entre enseignants et élèves et que les enseignants parviennent plus facilement à solliciter la participation des élèves. Cette étude révèle que la communication productive et satisfaisante dans une classe doit mettre en relief les mêmes principes de réciprocité, de mutualité, d'accordage, de timing régulier et de tours de parole qui soutiennent la communication entre parents et bébés. Les émotions vives de la voix d'un enseignant, son langage et ses comportements non verbaux sont sans doute aussi importants que les faits intemporels et les exercices de pensée et d'habileté qu'il ou elle cherche à transmettre.

Ces règles relationnelles s'appliquent dans tous les contextes de l'apprentissage, depuis la crèche jusqu'à l'université. Elles opèrent également dans des contextes d'apprentissage spontané (Bruner, 1996 ; Rogoff et coll., 2003). Elles soutiennent l'effort individuel et la découverte de la réussite à travers une expérience d'absorption et de concentration dans la maîtrise de tâches complexes (Czikszentmihalyi, 1990).