

Développement d'un questionnaire d'évaluation subjective de la qualité de la voix et de la parole (QEV)

*Sangsue, J.**, Siegwart, H.**, Cosnier, J., Cornu, J. & Scherer, K. R.***

** FPSE, Université de Genève
40 bv du Pont d'Arve, 1205 Genève
Tél. : +41 22 705 92 07 - Fax : +41 22 705 92 19
e-mail : sangsue@fapse.unige.ch

RESUME

La voix représente un indicateur puissant de l'état affectif d'un locuteur et se montre extrêmement sensible aux variations de ces états. Les recherches réalisées jusqu'alors ont mené, grâce à des mesures quantitatives et objectives, à une connaissance meilleure des propriétés acoustiques de productions vocales émotionnelles. La présente étude se centre sur l'aspect d'évaluation qualitative de la voix. Notre objectif est de créer une grille standardisée permettant de qualifier le plus exactement possible la voix d'un locuteur, l'intérêt étant de parvenir à dégager des patterns consistants de qualités vocales spécifiques aux états émotionnels. Deux études pilotes sont présentées, retraçant le développement d'une échelle qualitative de la voix et les premiers résultats issus de son application.

ABSTRACT

Voice represents a powerful indicator of a speaker's affective state and is also very sensitive to the variations of these states. Recent empirical work has led to a better understanding of the acoustics correlates of the affective vocalizations, based on quantitative measurements. The present study provides a complement by focusing on the qualitative aspects of vocal communication. Our purpose is to create a standardized instrument that can be used by listeners to subjectively evaluate a speaker's vocal productions. Of major interest is the attempt to find consistent patterns of vocal productions corresponding to specific emotional states. Two pilot studies are presented explaining the development of the qualitative voice scale and the first results of its application. It is suggested that the use of such a tool represent a convenient complementary diagnostic instrument in clinical settings.

1. INTRODUCTION

Nul ne met plus en doute l'aptitude innée de l'homme à communiquer ses émotions. Les psychologues l'étudient depuis longtemps au travers des expressions faciales, des gestes et des postures. Les dernières décennies ont favorisé un essor de l'étude de la communication vocale des affects. La voix constitue un canal essentiel par lequel sont extériorisées les émotions.

La psychologie clinique, développementale, sociale et la psychiatrie représentent des domaines où les caractéristiques affectives exprimées vocalement constituent des indices utiles à la compréhension du comportement humain. Pourtant, il ne semble pas exister de consensus quant aux dimensions ou aux paramètres à étudier. D'aucuns considèrent essentiellement les mécanismes impliqués dans la production [Gre75; Sim80], d'autres privilégient l'étude du versant réceptif [Dav64; Roe76; Pit93]. La plupart s'accordent toutefois sur le fait que l'émotion, en faisant varier le tonus musculaire entre autres, influence le fonctionnement des systèmes respiratoire, phonatoire et articulaire qui, en réaction, modifient certaines dimensions acoustiques de la voix. Celles-ci seraient à l'origine de l'impression subjective que se forme l'auditeur de l'état affectif du locuteur. Phénomènes intra-organiques d'abord, les caractéristiques acoustiques interviennent également à un niveau inter-organique où elles participent à la fonction sociale et communicative de la voix. Au-delà de ce registre acoustique, il semble indispensable de tenir compte également d'un niveau psycho-acoustique pour une compréhension précise et complète des processus impliqués dans la transmission vocale des affects. C'est-à-dire qu'il convient d'envisager les versants de production et de réception en plus de leur fonction respective et d'en examiner les liens.

Le modèle des lentilles de Brunswik, auquel nous nous référons (voir Fig. 1), permet une étude systématique de ces trois pôles en évoquant trois étapes principales: encodage, transfert et décodage. L'émetteur produit un message (encodage); il y a transfert de ce message et le récepteur élabore une impression ou une représentation subjective de ce message (décodage). Lors de l'étape de production ou encodage, l'énoncé verbal du locuteur est codé sous forme de signes, les indices « distaux », objectivement mesurables au niveau acoustique. Au moment du décodage (réception), ces indices sont réélaborés par l'auditeur en indices « proximaux » dont la représentation permettra l'inférence de l'état émotionnel du locuteur.

A l'émission d'un message verbal correspond donc l'organisation de signes objectifs, les indices « distaux », et subjectifs, les indices « proximaux ». L'un et l'autre type d'indices doivent être considérés pour une évaluation rigoureuse du comportement vocal. Les données empiriques et théoriques disponibles relatives aux mesures objectives ou aspects quantitatifs du versant production ont permis une description des propriétés acoustiques de l'émotion au moyen de divers paramètres (temporaux, d'amplitude, de fréquence, ainsi que de paramètres spectraux) et ont mené à la spécification de patterns acoustiques propres aux émotions [pour une revue, voir Ban96; Hes91; Sch86]. On a pu en outre démontrer que certains états pathologiques s'accompagnent de caractéristiques vocales particulières [Sch81; Sch88; Sch89]. Par exemple, les patterns acoustiques de patients dépressifs semblent se définir par une voix monotone, une réduction de la variabilité de la Fo, une intensité abaissée, des imprécisions articulatoires et de longues pauses [EII96].

Le pôle réception et les aspects qualitatifs de la voix appréhendés par l'auditeur restent le parent pauvre des études sur l'expression vocale des affects. Néanmoins, certaines études ont permis de mettre en évidence que des sujets auditeurs naïfs peuvent reconnaître diverses émotions exprimées par la voix, en jugeant des échantillons de phrases dépourvues de sens [Ban96; Pit93; Sch91]. Si le récepteur réussit avec une certaine précision à reconnaître les émotions du locuteur, les indices distaux du locuteur devraient véhiculer des informations au travers des caractéristiques acoustiques du message codé [Sch86]. Dès lors, il s'agit de déterminer quels sont, pour l'auditeur, les traits vocaux importants pour la discrimination des émotions d'un locuteur sur la base d'un énoncé vocal [Sch82].

Nous avons développé un questionnaire d'évaluation subjective visant la mise en correspondance des niveaux de production et de réception; celui-ci devrait permettre de caractériser les relations entre indices distaux et indices proximaux, et permettre à un auditeur « naïf » de définir à l'aide de vocables communément employés l'impression qu'il a des indices distaux qui lui sont transmis par un locuteur. Le questionnaire, ou grille d'évaluation, rédigé en langue française devrait être le plus exhaustif possible pour que tout type de voix puisse être caractérisé sur la base des termes proposés. En collaboration avec des professionnels de la voix (psycholinguistes, phoniâtres, phonéticiens)¹, un examen systématique de la littérature allemande et anglaise a été effectué pour regrouper les qualificatifs les plus

couramment usités pour qualifier la voix. Des analyses factorielles ont été réalisées de façon à regrouper les items et à en réduire le nombre. Les items sélectionnés ont été traduits en français puis disposés selon deux types d'échelles: unipolaire et bipolaire.

Cet article décrit les premières applications du questionnaire d'évaluation de la voix (QEV²) constitué via deux études de validation. Plus précisément, la première étude a pour but de tester le questionnaire QEV et de déterminer si les critères qualificatifs proposés sont appropriés pour définir la qualité d'une production vocale émotionnelle. La seconde étude vise à évaluer la validité de ces termes relativement à toute production vocale, c'est-à-dire non émotionnelle également.

2. ÉTUDE 1

MÉTHODE

Participants

17 étudiants en psychologie à l'Université de Genève, hommes et femmes, ont pris part à cette expérience. Tous étaient âgés de 18 à 25 ans.

Variables indépendantes

Dix échantillons de voix représentatifs de 5 émotions (peur, colère, dégoût, joie, tristesse) (pour une description des échantillons, voir plus bas).

Variables dépendantes

Évaluation subjective des échantillons vocaux, soit les réponses des sujets aux grilles A et B (Annexes 2 et 3).

Matériel

Grille d'évaluation A: Afin de déterminer si les échantillons vocaux sélectionnés caractérisent effectivement les émotions attendues, nous avons établi une première échelle (Grille A, Annexe 2) permettant de juger dans quelle proportion les 5 émotions (peur, colère, dégoût, joie, tristesse) sont contenues dans l'échantillon entendu.

Grille d'évaluation B: Du fait d'un glissement fréquent dans l'emploi habituel des termes parole et voix, nous avons pris en considération chacun de ces champs sémantiques pour fixer le choix des adjectifs de cette grille. En effet, la parole, définie en tant qu' instrument vocal d' interaction sociale, porteuse d'éléments linguistiques mais aussi d'informations non linguistiques, se réfère à un registre psycho-acoustique. La voix, quant à elle, porte davantage sur une dimension acoustique, puisqu' elle résulte de l'émission du son laryngé modifié par les cavités de résonance (pharynx, bouche et fosses nasales). La

¹ Dr. Boves, L. et Dr. Van Bezooigen, R. (Nimègue, Hollande) ainsi que Dr. Cornut, G., Prof. Cosnier, J., Duchêne, A., et Grosjean, M. ("Groupe de Recherches sur les Interactions Communicatives, Université Lyon 2, Lyon, France).

² Le questionnaire QEV peut être obtenu auprès des auteurs, ainsi que les Grilles A et B mentionnées Annexes 2 et 3.

grille présente, d'une part, 18 qualificatifs pour la voix et, d'autre part, 12 qualificatifs pour la parole, à évaluer sur des échelles unipolaires ou bipolaires. Les échelles bipolaires sont formées de deux adjectifs dont on a jugé qu'ils sont des oppositions sémantiques ou antonymes (e.g. faible-forte). Ces adjectifs sont liés par une échelle numérique continue, allant de 3 à 0 et de 0 à 3, 3 signifiant le degré de pertinence du qualificatif désigné, 0 signifiant l'absence de pertinence de l'un et l'autre adjectifs. Les échelles unipolaires sont formées d'un adjectif dont on n'a pas trouvé d'opposition sémantique (e.g. nasillarde). Par mesure de lisibilité dans la mise en page, cet adjectif est précédé d'une échelle allant de 0 à 3, 3 signifiant la pertinence et 0 signifiant au contraire l'absence de pertinence. La grille d'évaluation est divisée en trois parties: une première partie regroupe une série de 5 échelles bipolaires; elle est suivie d'une série de 8 items unipolaires; enfin, une 3ème série est constituée de 12 adjectifs mentionnés deux à deux sur 6 échelles bipolaires (Annexe 3).

Echantillons de voix: 10 échantillons de voix ont été enregistrés sur cassette audio. Ces échantillons sans signification sont constitués de suites de sons de différentes langues indo-européennes (d'une longueur de 7 syllabes). Elles sont récitées par des comédiens professionnels de la Radio Allemande (2 hommes et 2 femmes) et [Sch91, pour plus de détails].

Chaque échantillon vocal est présenté 4 fois à la suite, de façon à ce que le sujet continue à y avoir auditivement accès pendant qu'il remplit la grille d'évaluation.

Radio-Cassette stéréo OUC stéréo/OM : Nous disposons d'un radio-cassette de type «grand public », puissance max. 10 W, microphone intégré.

Procédure

Les participants étaient réunis dans une salle et assis derrière une grande table rectangulaire au milieu de laquelle était disposé le radio-cassette, haut-parleur orienté vers le plafond. Leur tâche était de déterminer d'abord, pour chacun des 10 échantillons de voix, dans quelle proportion les émotions de peur, colère, dégoût, joie et tristesse étaient incluses dans l'échantillon qu'ils entendaient. Ils répondaient au moyen de la Grille A et étaient tenus de donner 5 réponses pour chaque échantillon, soit une réponse par émotion proposée. L'expérimentateur fait ensuite écouter à nouveau quatre fois de suite, le premier échantillon vocal et chaque participant est alors invité à cocher sur la Grille B le degré de pertinence des adjectifs qualifiant l'échantillon de voix entendue. Dans le cas des échelles bipolaires, on prie le participant de ne cocher qu'un seul chiffre. L'expérimentateur attend que toutes les personnes aient complété la grille avant de passer à l'échantillon suivant. La procédure est répétée pour les 10 échantillons de voix. Lorsque l'expérience est terminée, l'expérimentateur récolte les 170 grilles.

3. RÉSULTATS

Les pourcentages de reconnaissance des émotions, calculés à partir de la Grille A, indiquent une correspondance importante entre les émotions souhaitées être inférées par nos échantillons vocaux et les émotions perçues. Une réserve s'impose toutefois concernant les échantillons sensés représenter l'émotion de joie qui ne sont que faiblement identifiés comme exprimant de la joie. Nous les avons néanmoins maintenus pour la poursuite de notre objectif, de façon à maintenir une certaine diversité de notre matériel et ne pas avoir que des échantillons vocaux d'émotions négatives.

Quant aux données issues de la Grille B, afin de déterminer la pertinence des qualificatifs proposés, nous avons établi un calcul de la fréquence d'apparition du 0 en pourcentage (l'apparition d'un 0 signifiant la non-pertinence des qualificatifs). Les résultats sont donnés dans la Table 1.

Table 1: Pourcentages de zéro (càd. adjectif non pertinent)

Adjectifs proposés	% de 0 <50%	% de 0 >50%
Faible/fort	13	
Grave/aiguë	8	
Claire/sourde	8	
Ferme/tremblante	8	
Fluette/ample	24	
Rugueuse		60
Accompagnée de souffle	37	
Dure		53
Chaude	37	
Rauque		59
Douce	37	
Nasillarde		71
Perçante		62
Lente/rapide	15	
Nette/floue	6	
Hésitante/aisée	13	
Monotone/modulée	12	
Fluide/hachée	10	
Accentuée +/-	16	

De nettes différences en fonction du type d'échelle peuvent être relevées. Pour les échelles bi-polaires, tous les critères proposés obtiennent une fréquence de non-pertinence inférieure à 20%. En revanche, la majorité des adjectifs proposés en unipolaire remportent des pourcentages de non-pertinence excédant 50%. En effet, si les échelles bipolaires laissent supposer que leurs constituants sont appropriés pour la caractérisation subjective des énoncés vocaux, les adjectifs des échelles unipolaires apparaissent pour leur part moins représentatifs. L'interprétation de nos données devra rendre compte

de ces différences et nous y reviendrons dans notre partie discussion. La conclusion essentielle qui doit être retenue est que les termes qualificatifs proposés, en l'occurrence via les échelles bipolaires, semblent bien adaptés pour l'élaboration d'un système standardisé d'évaluation subjective de la voix.

4. ÉTUDE 2

MÉTHODE

Participants

16 femmes et 8 hommes âgés de 20 à 40 ans, de cultures différentes, étudiants ou de profession libérale ont participé à ce deuxième pilotage.

Variables indépendantes

Trente quatre échantillons de voix, masculines et féminines, d'une durée moyenne de 30 secondes.

Variables dépendantes

Réponses des sujets au QEV (Annexe 1).

Matériel

Grille d'évaluation: la Grille B utilisée dans le cadre de l'étude 1 a été reprise, avec une modification apportée concernant les items «dure » et «douce ». Ces derniers ont été mentionnés sur une échelle bipolaire et non pas sur deux échelles indépendantes, unipolaires. Cette décision a été motivée par le fait que ces deux termes sont antonymiques. Notre grille reste divisée en trois parties, respectivement de 2 x 6 échelles bipolaires et de 1 x 6 échelles unipolaires.

Echantillons de voix: 34 échantillons de voix masculines et féminines relatant un même texte standard (de contenu sémantique neutre). La durée des échantillons variait quelque peu en fonction du lecteur (durée moyenne d'environ 30 secondes). Parmi les personnes lisant le texte, trois d'entre elles le racontaient en langue étrangère (allemand, anglais et japonais) qui étaient les langues d'origine des locuteurs. Chaque échantillon est présenté quatre fois de suite comme dans la première procédure.

Radio-Cassette stéréo OUC stéréo/OM: Des écouteurs étaient raccordés pour chaque sujet à un radio-cassette de type « grand public », microphone intégré, ce qui permettait des conditions d'écoute comparables.

Procédure

Les participants ont été divisés en trois groupes. Les passations ont été des passations collectives de 8 personnes. Leur tâche était de cocher, sur chaque grille, à quel degré le qualificatif évoqué leur paraissait pertinent pour caractériser la voix qu'ils entendaient. Ils devaient répondre à tous les items de la grille. Les participants étaient disposés autour d'une table et avaient à leur disposition 34 grilles d'évaluation. Chacun disposait également d'un feuillet de consignes dont il prenait connaissance. Lorsque l'expérience débutait, on passait un

premier échantillon de voix. L'étudiant(e) écoutait en ayant la possibilité de cocher ses réponses. A la fin de l'échantillon sonore, l'expérimentateur attendait que chaque participant ait terminé de remplir sa grille. Puis il leur faisait entendre le stimulus sonore suivant et ainsi de suite. Une pause était respectée à la dix-septième voix afin de permettre aux sujets de se "reposer". L'expérimentateur ramassait ensuite l'ensemble des grilles d'évaluation remplies par les participants ($N = 24 \times 34 = 816$).

5. RÉSULTATS

Etant donné les trois passations réalisées en groupe de 8 personnes par trois expérimentateurs différents, l'analyse de nos résultats nécessitait en pré-requis de s'assurer que ces groupes étaient comparables. Une analyse de variance a été effectuée par groupe d'items et pour chaque expérimentateur. Un effet significatif du groupe est apparu ($F(2, 813) = 49.78, p < .001$), particulièrement sur les items unipolaires, indiquant dans un des groupes une fréquence élevée dans le pourcentage de réponses 0. Ce groupe de 8 personnes a été remplacé par 10 autres participants, ce qui porta l'effectif de participants à 26. Une nouvelle analyse montre qu'il n'y a plus de différence significative entre les 3 groupes ($F(2, 881) = 2.64, p < .07$).

Une différence significative dans le pourcentage de réponses "non-pertinentes" pour le premier échantillon de voix ainsi que pour ceux de voix étrangère en comparaison à l'ensemble des autres stimuli a également pu être mise en évidence. Ces échantillons ont été supprimés pour l'analyse de nos données qui a été menée sur les réponses aux 30 stimuli vocaux restants.

Etant donné le nombre d'échantillons vocaux et leur répétition, la passation durait approximativement 60 minutes, pause y compris après le 17ème échantillon vocal. Une analyse de variance a été réalisée par groupes d'items sur la première partie (échantillons 1 à 17) et la seconde partie (échantillons 18 à 34) des voix, de façon à tester l'existence d'un éventuel effet de fatigue. Cette analyse ne révéla aucune différence significative entre les deux séries d'échantillons, ni pour les items bipolaires ni pour les items unipolaires ($p < .50$ pour les 6 premiers items bipolaires; $p < .90$ pour les 6 items unipolaires et $p < .50$ pour les 6 derniers items bipolaires)

Fréquence d'apparition du zéro (%)

Un calcul de la fréquence d'apparition du zéro a été effectué et traduit en pourcentages afin d'évaluer la pertinence des différents qualificatifs pour la caractérisation des échantillons de voix-stimuli. Il apparaît de nettes différences entre les trois parties d'items. Les résultats indiquent que les qualificatifs proposés pour caractériser la voix sont dans

l'ensemble bien adaptés puisqu'une majorité (soit les 2/3) des items obtiennent un score de 0 faible, inférieur à 20%. Plus important, les résultats indiquent à nouveau une nette différence entre les items bipolaires (dont le pourcentage de 0 est inférieur à 20) et les items unipolaires, remportant des pourcentages de 0 allant parfois au-delà de 50% (Table 2).

Table 2: Pourcentages de zéro (càd. adjectif non pertinent)

Adjectifs proposés	% de 0 <50%	% de 0 >50%
Items bipolaires		
Faible/fort	11	
Grave/aiguë	10	
Claire/sourde	5	
Ferme/tremblante	11	
Dure/Douce	21	
Lente/rapide	7	
Nette/floue	6	
Hésitante/aisée	7	
Monotone/modulée	6	
Fluide/hachée	8	
Accentuée ++/-	18	
Items unipolaires		
Rugueuse		58
Accompagnée de souffle	41	
Chaude	38	
Rauque		57
Nasillarde		70
Perçante		64

Afin de préciser davantage la relation existant entre les deux types d'échelles, des corrélations (sur les zéro) ont été effectuées entre les deux groupes d'items bipolaires et le groupe d'items unipolaires (groupes nommés respectivement A (= 6 premiers items bipolaires), C (= 6 items unipolaires) et B (= 6 derniers items bipolaires). La corrélation entre A et C de $r = .58$, $p < .001$, la corrélation entre A et B de $r = -.08$, $p < .32$, et la corrélation entre B et C de $r = -.22$, $p < .10$. La corrélation significative entre A et C confirme une dépendance entre les deux groupes d'échelles bipolaires: un taux de 0 élevé dans l'échelle A est directement lié à un taux de 0 élevé dans l'échelle C et vice versa. Au contraire, les corrélations entre les échelles bipolaires et le type unipolaire indiquent qu'il n'y a pas de lien entre le fait d'attribuer des 0 aux échelles bipolaires et des 0 aux adjectifs unipolaires. Il semblerait que les deux types d'échelles aient des statuts différents et qu'elles relèveraient d'un traitement indépendant. Deux propositions d'interprétation de ces différences en terme

de contenu et de forme des échelles seront abordées dans notre partie discussion et conclusion.

Analyse factorielle

Intéressés à savoir quelles caractéristiques de la voix et de la parole covariaient, une analyse factorielle a été conduite. Celle-ci a permis de dégager 5 facteurs (Table 3), expliquant ensemble 65 % de la variance totale, libellés comme suit: « Rapport entre son périodique et son non-périodique »; « Qualité de la parole »; « Aisance de la parole »; « Hauteur de la voix »; « Courbe d'intonation de la parole ». Deux items ("grave-aiguë" et "ferme-tremblante") sont moins corrélés dans chaque facteur. En les retirant de l'analyse, un maximum de fiabilité pour les facteurs est obtenu et davantage de consistance interne, à savoir plus d'homogénéité entre les items formant les facteurs.

Table 3: Coefficients de saturation obtenus pour l'analyse factorielle menée sur les caractéristiques vocales.

	1 ^a	2	3	4	5
	23	13.2	11.6	7.8	6.3
	%	%	%	%	%
Rauque	.80	-.10	-.22	.01	.05
Rugueuse	.80	-.10	-.28	.02	.05
Claire-sourde	.61	-.36	.10	-.23	-.15
Accompagnée de souffle	.60	.09	.39	.13	-.04
Hésitante-aisée	-.15	.81	-.03	-.08	.19
Lente-rapide	.09	.73	-.03	.24	-.11
Fluide-hachée	.21	-.70	.10	.11	-.18
Nette-floue	.44	-.52	.23	-.02	-.16
Fluette-ample	.17	.06	-.78	-.15	.08
Faible-forte	-.03	.26	-.65	.26	.14
Ferme-trembl.	.45	-.42	.46	.00	-.09
Perçante	-.19	.20	-.05	.72	.06
Dure-douce	-.23	-.03	.37	-.62	.38
Nasillarde	.15	-.22	.17	.54	.11
Grave-aiguë	-.42	.26	.45	.47	.09
Accentuée ++/--	-.01	.01	.11	-.22	-.77
Monotone-modulée	-.10	.30	-.08	-.04	.74
Chaude	.09	.09	-.05	-.51	.61

Note. ^a le pourcentage indique la part de variance expliquée par le facteur.

6. DISCUSSION

Ces premiers résultats démontrent la « faisabilité » d'un système standardisé destiné à l'analyse de la qualité de la voix (représentative d'émotion ou non). Une fiabilité statistique satisfaisante a été trouvée pour la plupart des items d'évaluation vocale. La majorité de nos qualificatifs semblent convenir pour refléter l'impression subjective d'un récepteur exposé à

l'énoncé d'un message verbal. Dans plus de 80% des cas, les adjectifs ont été décrits comme pertinents.

Certains cependant se sont révélés inadaptés et ont soulevé différentes questions quant au statut des échelles employées dans le questionnaire (bipolaire vs. unipolaire). En effet, les termes jugés non pertinents pour caractériser la voix sont mentionnés par l'échelle unipolaire. Nos analyses ont confirmé une indépendance entre l'utilisation des deux types d'échelles (bipolaires/ unipolaires). Nous ne pouvons pas tirer de conclusions ou d'interprétations tangibles sur le statut respectif de ces échelles, car nous ne sommes pas en mesure de déterminer, sur la base de nos données, si le problème lié à l'utilisation de l'échelle unipolaire a trait à sa forme ou à son contenu.

Une proposition serait que les échelles unipolaires laissent la possibilité au sujet d'évincer clairement le qualificatif proposé. Dans les échelles bipolaires la possibilité existait aussi en cochant le chiffre 0. Mais les participants ont peut-être mal interprété le zéro et considéré ce dernier comme étant l'un et l'autre dans la même proportion. Ceci forçait du même coup le participant à davantage de précision et à opter un peu plus pour l'un ou l'autre, les adjectifs étant hautement contradictoires. Il se pourrait également que la bipolarité renforce le champ de signification des items en question et facilite la tâche des sujets. Enfin, il n'est pas exclu que les adjectifs présentés en unipolarité ne soient pas simplement inappropriés pour caractériser la voix et les émotions présentées. En effet, accompagnée de souffle, rugueuse, par exemple, ne nous semblent a posteriori pas être des vocables « populaires » et couramment employés par les individus.

Les analyses factorielles menées en deuxième partie ont défini des facteurs à propos desquels il est intéressant de constater que leurs constituants appartiennent autant au registre psycho-acoustique qu'au registre acoustique. Par exemple, le facteur 4= hauteur de la voix, défini par perçante, dure-douce, nasillarde, grave-aigüe, comporte deux critères appartenant à la dimension acoustique (dure-douce, grave-aigüe) et deux critères relatifs à la dimension psycho-acoustique (perçante, nasillarde). Ceci accrédite l'utilité de prendre en considération les relations entre pôles production (registre purement acoustique: la voix) et réception (registre psycho-acoustique: la parole) pour une compréhension des aspects de transmission vocale des affects.

Cette étude, de type exploratoire, comporte un certain nombre de limites; de ce fait, elle mériterait à nos yeux d'être développée. Après les traitements liminaires que nous venons de présenter, il s'avère nécessaire de poursuivre la validation de cette grille d'évaluation, afin de vérifier et modifier en conséquence sa structure, voire son contenu. Notamment, il conviendrait d'approfondir davantage la question des échelles concernant le statut du zéro et/ou la pertinence du qualificatif proposé. L'intérêt serait d'aboutir à la validation d'un système standardisé applicable à des états émotionnels et non-émotionnels de façon à pouvoir

dégager des patterns consistants quant aux caractéristiques acoustiques qu'ils évoquent subjectivement chez l'auditeur naïf ou encore chez un codeur formé. Comme il est actuellement tenu pour acquis qu'un grand nombre de troubles émotionnels ou de pathologies ont des retentissements sur les paramètres acoustiques de la voix, un tel instrument pourrait constituer un outil de diagnostique pour la psychologie clinique. Contrairement aux mots, les caractéristiques vocales échappent aux distorsions introduites parfois en fonction d'une désirabilité sociale ou d'une volonté de présentation de soi positive.

La psychologie sociale pourrait y trouver un bénéfice puisque la grille serait susceptible de qualifier l'importance relative de certaines caractéristiques vocales dans la formation d'impressions et de jugements sur le locuteur. Ce qui ne semble pas sans valeur pour comprendre certains phénomènes sociaux (e.g. les processus d'attribution, dont on connaît l'influence sur nos comportements, nos perceptions d'autrui et donc les interactions sociales).

Nous encourageons la poursuite de notre objectif majeur tendant à la mise au point un questionnaire valide qui permettant de préciser spécifiquement (1) sur quels traits vocaux le récepteur se base pour interpréter l'état émotionnel d'un locuteur, (2) si certaines émotions sont plus facilement discriminables que d'autres et (3) s'il existe des patterns de critères qualificatifs typiques aux états affectifs.

Bibliographie

- [Gre75] Green, R. S., & Cliff, N. (1975). Multidimensional comparisons of structures of vocally and facially expressed emotion. Perception & Psychophysics, 17(5), 429-438.
- [Sim80] Simonov, P. V., Frolov, M. V., & Ivanov, E. A. (1980). Psychophysiological Monitoring of Operator's Emotional Stress in Aviation and Astronautics. Aviation, Space and Environmental Medicine, January, 46-49.
- [Dav64] Davitz, J. R. (1964). Auditory correlates of vocal expressions of emotional meaning. In J.R. Davitz (Ed.), The communication of emotional meaning (pp 101-112). New York: Mac Graw-Hill.
- [Roe76] Roessler, R., & Lester, J.W. (1976). Voice Predicts Affect during Psychotherapy. The Journal of Nervous and Mental Disease, 163(3), 166-176.
- [Pit93] Pittam, J., & Scherer, K. R. (1993). Vocal Expression and Communication of Emotion. In M. Lewis & J. M. Haviland (Eds), Handbook of Emotions (pp. 185-198). London: The Guilford Press.

- [Ban96] Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology, 70(3), 614-636.
- [Hes91] Hess, U., Kappas, A., & Scherer, K. R. (1991). Multichannel communication of emotion: synthetic signal production. In K. R. Scherer (Ed.), Facets of Emotion: Recent Research (pp. 161-182). Hillsdale, NJ: Lawrence Erlbaum Associates.
- [Sch86] Scherer, K. (1986). Vocal Affect Expression: A review and a Model for Future Research. Psychological Bulletin, 99(2), 143-165.
- [Sch81] Scherer, K. R. (1981). Vocal indicators of stress. In J. K. Darby (Ed.), Speech evaluation in psychiatry (pp. 171-187). New York: Grune & Stratton.
- [Sch88] Scherer, K. R., & Zei, B. (1988). Vocal Indicators of Affective Disorders. Psychother Psychosom, 49, 179-186.
- [Sch89] Scherer, K. R., & Zei, B. (1989). La Voix comme Indice Affectif. Revue Médicale de la Suisse Romande, 109, 61-66.
- [Ell96] Ellgring, H. & Scherer, K. R. (1996). Vocal indicators of mood change in depression. Journal of Nonverbal Behavior, 20, 83-110.
- [Sch82] Scherer, K. (1982). Methods of Research on vocal communication: Paradigms and Parameters. In K. R. Scherer & P. Ekman (Eds), Handbook of methodes in non-verbal behavior research (pp. 136-198). Cambridge, Paris: Cambridge University Press, Editions de la Maison des Sciences de l'Homme.
- [Sch91] Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal Cues in Emotion Encoding and Decoding. Motivation and Emotion, 15(2), 123-148.

Figure

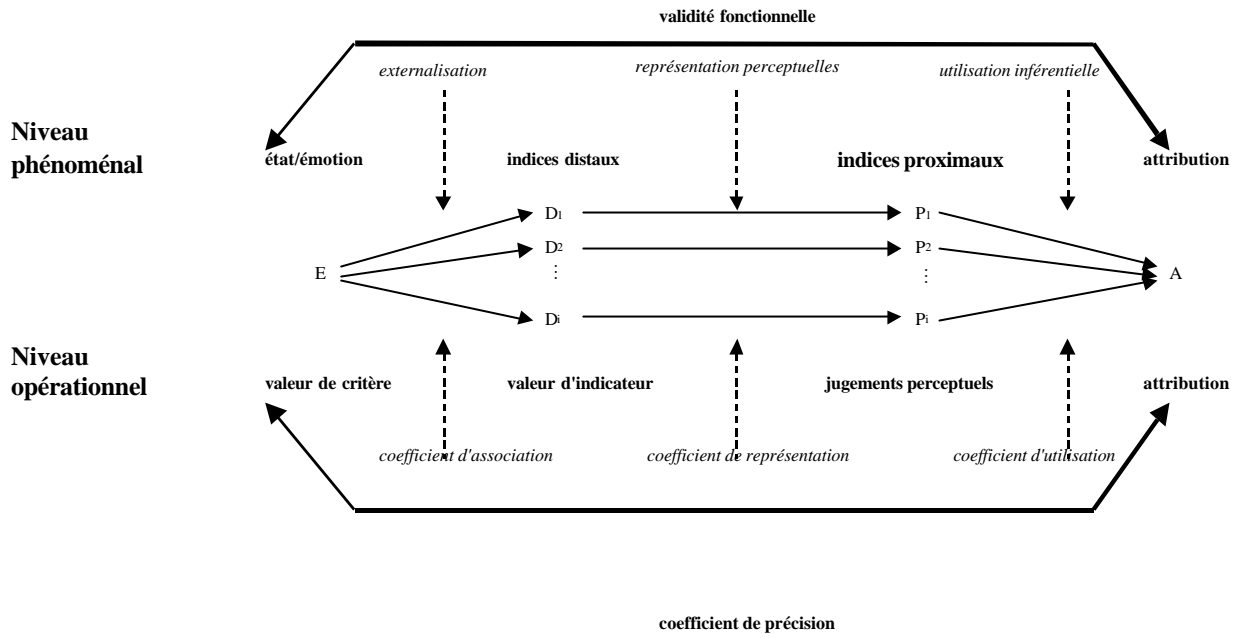


Figure 1: Adaptation du modèle "lens" de Brunswik pour la perception de l'émotion dans la voix [Sch86].