Segmentation du contour intérieur des lèvres en combinant contours actifs et modèles paramétriques

Sébastien Stillittano[1]

[1]Vesalis
Clermont-Ferrand, France

sebastien.stillittano@lis.inpg.fr

Alice Caplier[2]

[2]GIPSA-lab/DIS
Grenoble, France

alice.caplier@inpg.fr

Résumé

Les applications de lecture labiale requièrent des informations précises sur le mouvement et la forme des lèvres, caractérisées à la fois par le contour extérieur et par le contour intérieur de la bouche. Dans cet article, nous introduisons une nouvelle méthode de détection du contour intérieur. A partir du contour extérieur donné par un algorithme préexistant, nous utilisons des points clefs pour initialiser un contour actif appelé "jumping snake". Grâce à une information optimale de gradients de luminance et de chrominance, le contour actif ajuste la position de 2 modèles paramétriques différents composés de cubiques (un premier modèle pour des bouches fermées et un second pour des bouches ouvertes). Les modèles donnent un contour intérieur flexible et précis. Finalement, nous présentons plusieurs résultats expérimentaux démontrant l'efficacité de l'algorithme proposé.

Mots clefs

Contours actifs (« jumping snake »), modèle paramétrique, segmentation, lèvres, contour intérieur.

1 Introduction

De nombreuses études ont montré que l'information visuelle peut améliorer sensiblement la compréhension de la parole en environnement bruité [1] [2]. Le mouvement et la forme des contours intérieur et extérieur des lèvres donnent des informations utiles aux applications de lip reading. Ainsi, de nombreuses recherches ont été effectuées pour obtenir le contour extérieur, mais peu d'études portent sur l'extraction du contour intérieur; la raison principale étant que celui-ci est plus difficile à extraire du fait du contenu non uniforme de l'intérieur de la bouche. En effet, il peut y avoir différentes configurations lors d'une conversation. A l'intérieur de la bouche, il existe des zones qui peuvent avoir le même aspect que les lèvres (gencives et langue), des zones brillantes (dents), ainsi que des zones très sombres (cavité orale). Chaque zone pouvant apparaître et disparaître continuellement lors d'une conversation.

Parmi les approches existantes pour l'extraction du contour intérieur, certaines s'appuient sur un modèle paramétrique composé d'un assortiment de courbes. Dans [3], Zhang utilise des modèles déformables pour la détection des contours intérieur et extérieur. Les modèles choisis sont formés de 3 ou 4 paraboles, selon que la bouche est fermée ou ouverte. La première étape est une estimation des candidats possibles pour les paraboles en analysant l'information de luminance. Ensuite, le bon modèle est choisi parmi le nombre de candidats trouvés, les informations de luminance et de couleur permettant d'ajuster le modèle. Cette méthode donne des résultats qui ne sont pas assez précis pour faire du lip reading, à cause de la simplicité et de la symétrie supposée du modèle associé à la bouche.

Dans [4], Beaumesnil et al. utilisent en premier lieu, 2 contours actifs pour l'extraction des contours intérieur et extérieur. L'étape suivante implique un modèle 3D du visage afin d'extraire des paramètres plus précis et d'affiner les résultats. Un algorithme de classification "k-means", basé sur une teinte non-linéaire caractérise trois classes : lèvre, visage et fond. A partir de cette classification, une boîte englobant la bouche est définie, et les points du contour actif extérieur sont initialisés sur 2 cubiques calculées à partir de la boîte. Les forces utilisées pour la convergence du snake combinent les informations de teinte non-linéaire et de luminance. Ensuite, un contour actif intérieur est initialisé sur le contour extérieur, puis réduit par un changement d'échelle anisotropique, en considérant la position du centre de la bouche et l'épaisseur des lèvres. L'inconvénient est que le snake doit être initialisé très proche du contour car celui-ci converge vers le premier minimum local de gradient. Or différents minima de gradient sont générés par la présence des dents ou de la langue, et peuvent amener à une mauvaise convergence. Dans [4], le modèle 3D du visage permet de corriger des mauvaises convergences, mais le clone ne donne pas de résultats assez précis.

Des méthodes statistiques peuvent aussi être utilisées pour segmenter des lèvres. Dans [5] et [6], Cootes et al. développent des modèles actifs statistiques pour la forme (AMS) et l'apparence (AAM). Forme et apparence d'un objet sont apprises à partir d'un assortiment d'images annotées manuellement. Ensuite, une analyse en composante principale (ACP) est exécutée pour obtenir les principaux modes de variation. Le modèle est ajusté

itérativement pour réduire la différence entre modèle et contour réel en utilisant une fonction coût. Dans [7], Luettin et al. construisent un AMS et dans [8], Gacon et al. construisent un AMS et un AAM pour la détection des contours des lèvres. L'intérêt principal est que les résultats sont très réalistes, mais l'entraînement des données doit traiter de tous les cas possibles de forme de bouche.

Le but de nos travaux est d'obtenir une segmentation précise du contour intérieur des lèvres pour des applications de lip reading. Nous avons développé un algorithme basé sur des contours actifs et des modèles paramétriques; les modèles représentent la forme a priori de la bouche et le "jumping snake", introduit dans [9], ajuste leurs positions.
Pour la segmentation du contour extérieur, nous utilisons l'algorithme proposé dans [9]. A partir du contour extérieur obtenu, nous détectons des points clefs, puis nous faisons converger 2 "jumping snakes" et définissons 2 modèles paramétriques différents (selon que la bouche est fermée ou ouverte) pour extraire le contour intérieur.

L'article est organisé de la manière suivante. Dans la section 2, nous décrivons brièvement l'extraction du contour extérieur proposé dans [9]. Les sections 3 et 4 expliquent comment est trouvé le contour intérieur selon que la bouche est fermée ou ouverte. Des résultats expérimentaux sont présentés en section 5. Finalement, la section 6 conclut cet article.

2 Extraction du contour extérieur

Dans [9], Eveno et al. introduisent un modèle paramétrique flexible composé d'une ligne brisée et de 4 cubiques pour décrire le contour extérieur des lèvres (voir fig. 1). Le modèle est initialisé par 6 points clefs et est ajusté en utilisant des informations de gradient calculées à partir de la pseudo-teinte [10] et de la luminance. Les points P_2, P_3 et P_4, reliés par une ligne brisée, forment l'arc de Cupidon, le point P_6 est le point le plus bas du contour et les points P_1 et P_5 représentent les coins de la bouche. 4 cubiques (γ_i), reliant P_2 et P_6 (resp. P_4 et P_6) à P_1 (resp. P_5), complètent le contour extérieur.

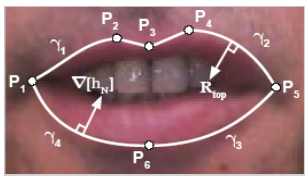

Fig. 1 - *Les points clefs et le modèle paramétrique [9]*

Notre algorithme pour la détection du contour intérieur s'inspire de l'algorithme décrit dans [9]. Notre méthode suppose tout d'abord que le contour extérieur a été bien segmenté et que nous pouvons utiliser les points clefs de ce contour pour initialiser notre algorithme. De plus, nous faisons l'hypothèse que les contours intérieur et extérieur sont reliés par les coins de la bouche (P_1 et P_5).
Nous développons 2 stratégies différentes et 2 modèles différents selon que la bouche est fermée ou ouverte.

3 Cas d'une bouche fermée

3.1 Modèle choisi

Le modèle paramétrique pour le contour intérieur, lorsque la bouche est fermée, est composé de 2 cubiques (γ_5 et γ_6) et d'une ligne brisée (voir fig. 4). La ligne brisée reliant les points P'_2, P'_3 et P'_4 du modèle permet de représenter la distorsion du contour intérieur due à l'arc de Cupidon et les deux cubiques, entre le point P'_2 (resp. P'_4) et le coin de la bouche P_1 (resp. P_5), complètent le contour.
Une étude expérimentale a montré qu'une parabole n'est pas assez précise pour décrire le contour intérieur. Les applications de lip reading demandent un contour intérieur précis, et, ce que nous pouvons appeler "arc de Cupidon intérieur", ne peut être représenté par une simple parabole reliant les coins de la bouche.

3.2 Initialisation du modèle

Pour une bouche fermée, le contour intérieur peut être vu comme une ligne sombre reliant les coins de la bouche. Nous utilisons la ligne L_{min} pour initialiser la recherche du contour. L_{min} est composée des pixels les plus sombres et de plus, les coins de la bouche ont été choisis sur cette ligne dans [9]. L_{min} est initialisée sur le pixel le plus sombre du segment $[P_3P_6]$ et grandit en ajoutant des pixels dans les 2 directions, gauche et droite. Pour chaque direction, seuls les 3 plus proches pixels sont candidats et le pixel ayant la luminance minimale est choisi. Comme le montre la fig. 2, L_{min} est déjà une bonne représentation du contour intérieur de la bouche.

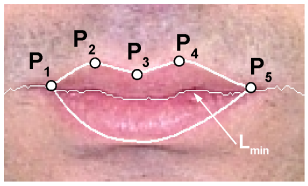

Fig. 2 - *Détection de L_{min}*

L_{min} est échantillonnée et donne le contour initial appelé C_1 (voir fig. 3). Nous trouvons 3 points P'_2, P'_3 et P'_4, afin de fixer les limites des 3 composantes de notre modèle. P'_3 se trouve sur le contour C_1, c'est le point le plus proche de la verticale passant par P_3. P'_2 est le point le plus haut du contour C_1 limité par les deux verticales passant par P_2 et P_3 (intervalle I_1 sur la figure 3). P'_4 est le point le plus haut du contour C_1 limité par les deux verticales passant par P_3 et P_4 (intervalle I_2 sur la figure 3). Les coins de la bouche sont les points P_1 et P_5 trouvés lors de la détection du contour extérieur.

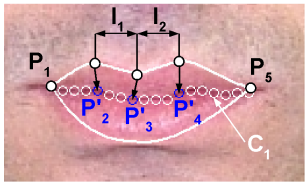

Fig. 3 - *Contour initial C_1 et détection des points clefs*

3.3 Optimisation du modèle

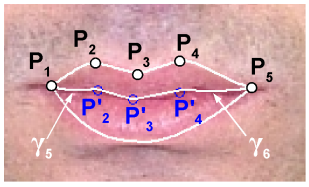

Fig. 4 – *Modèle paramétrique pour une bouche fermée*

A partir des points clefs détectés, le contour intérieur final est donné par la ligne brisée reliant P'_2, P'_3 et P'_4, et les 2 cubiques, entre le coin de la bouche P_1 (resp.P_5) et le point P'_2 (resp. P'_4). Les 2 courbes sont calculées par la méthode des moindres carrés.

4 Cas d'une bouche ouverte

La détection du contour intérieur est plus difficile lorsque la bouche est ouverte, à cause des variations non-linéaires d'apparence à l'intérieur de la bouche. En effet, lors d'une conversation, la zone située entre les lèvres peut prendre différentes configurations: 1) Dents , 2) Cavité orale, 3) Gencives et 4) Langue.

4.1 Modèle choisi

Le modèle paramétrique pour le contour intérieur, lorsque la bouche est ouverte, est composé de 4 cubiques (voir fig. 7). Pour une bouche ouverte, "l'arc de Cupidon intérieur" (voir section 3.1), est moins prononcé que pour une bouche fermée; ainsi 2 cubiques suffisent pour précisément extraire le contour intérieur supérieur des lèvres. Avec 4 cubiques, le modèle est flexible et permet de surmonter le problème de la segmentation du contour intérieur pour des bouches asymétriques.

4.2 Initialisation du modèle

Deux contours actifs appelés "jumping snakes", comme introduit dans [9], sont utilisés pour ajuster le modèle; un 1er pour le contour supérieur et un 2nd pour le contour inférieur.

La convergence d'un "jumping snake" est une succession de phases de croissance et de saut. Le snake est initialisé à partir d'un germe, puis, il grandit en ajoutant des points à gauche et à droite du germe. Chaque nouveau point est trouvé en maximisant un flux de gradient à travers le segment formé par le point courant à ajouter et le point précédent. Finalement, le germe saute vers une nouvelle position plus proche du contour recherché. Les processus de croissance et de saut sont répétés jusqu'à ce que l'amplitude du saut soit inférieure à un certain seuil. L'initialisation de nos 2 snakes commence par la recherche de 2 points (P_7 et P_8 de la fig. 5) sur les contours supérieur et inférieur, et appartenant à la verticale passant par P_3. La difficulté de la tâche réside dans le fait que nous pouvons trouver différentes zones entre les lèvres, qui peuvent avoir des caractéristiques (couleur, texture ou luminance) similaires ou complètement différentes que celles des lèvres, quand la bouche est ouverte. L'objectif est de trouver l'information pertinente qui peut accentuer le contour intérieur dans toutes les configurations possibles. Une étude expérimentale sur des milliers d'images de visage a montré qu'aucun espace ne peut atteindre ce but et nous devons considérer une combinaison entre des informations venant de différents espaces; chacune des informations accentuant le contour pour une des configurations. Par exemple, les lèvres sont caractérisées par une forte pseudo-teinte et une forte composante rouge, les dents sont brillantes et saturées en couleur, la cavité orale est très sombre, alors que les gencives et la langue peuvent avoir le même aspect que les lèvres. Nous avons construit expérimentalement 2 gradients (G_1 et G_2, voir eq.1 et 2) combinant différentes informations pour trouver P_7 et P_8.

P_7 est trouvé en cherchant le maximum du gradient G_1 entre P_3 et P_6. P_8 est trouvé en cherchant le maximum du gradient G_2 entre P_3 et P_7. Afin d'éviter les mauvaises détections dues au bruit, nous accumulons les différents gradients sur une largeur de 10 pixels autour de P_3 et nous choisissons le plus fort gradient cumulé.

$$G_1(x,y) = \nabla\left[Cr_N(x,y) + h_N(x,y) + L_N(x,y)\right] \quad \textbf{(1)}$$

$$G_2(x,y) = \nabla\left[L_N(x,y) - Cr_N(x,y) - S_N(x,y) - 3*h_N(x,y)\right] \quad \textbf{(2)}$$

Où Cr_N vient de l'espace YCbCr, h_N est la pseudo-teinte, L_N est la luminance et S_N est la saturation de l'espace HSV, normalisées entre 0 et 1. La pseudo-teinte, introduite dans [10], est h = R/R+G, où R et G viennent de l'espace couleur RGB. La pseudo-teinte accentue la différence de contraste ente les lèvres et la peau [11].

Des points P_7 et P_8, nous calculons 2 germes P'_7 et P'_8 pour l'initialisation des "jumping snakes". P'_8 représente les ¾ du segment $[P_3P_8]$ et P'_7, les ¾ du segment $[P_6P_7]$ (voir fig. 5). De ce fait, les germes sont plus proches des contours intérieurs que d'éventuels contours bruités.

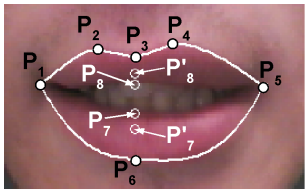

Fig. 5 - *Détection des germes pour les snakes*

Pour la convergence des snakes, nous devons aussi trouver des gradients qui accentuent le contour intérieur dans chaque configuration. De la même façon, nous avons construit expérimentalement 2 combinaisons de plans.
Pour le contour supérieur, la convergence du 1er snake donne le contour initial C_2. P'_8 est le germe et les paramètres sont choisis de manière à ce que les branches du snake aient tendance à descendre. G_3 (voir eq.3) est le gradient utilisé pour la phase de croissance du snake.
Pour le contour inférieur, la convergence du 2nd snake donne le contour initial C_3. P'_7 est le germe et les paramètres sont choisis de manière à ce que les branches du snake aient tendance à monter. G_4 (voir eq.4) est le gradient utilisé pour la phase de croissance (voir fig. 6).

$$G_3(x,y) = \nabla\left[R_N(x,y) - u_N(x,y) - h_N(x,y)\right] \quad \textbf{(3)}$$

$$G_4(x,y) = \nabla\left[L_N(x,y) + u_N(x,y) + h_N(x,y)\right] \quad \textbf{(4)}$$

Où R_N est la composante rouge de l'espace RGB, L_N est la luminance, u_N vient de l'espace CIELuv [12] et h_N est la pseudo-teinte, normalisées entre 0 et 1.

Nous prenons les 2 points les plus proches (P''_8 et P''_7) de la verticale passant par P_3 sur les contours C_2 et C_3 comme points clefs pour notre modèle.

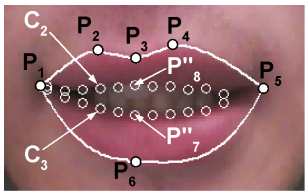

Fig 6 - *Convergence des "jumping snakes"* .

4.3 Optimisation du modèle

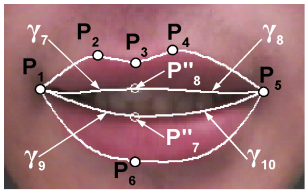

Fig. 7 – *Modèle paramétrique pour une bouche ouverte*

A partir des points clefs détectés, le contour intérieur final est donné par 4 cubiques entre P_1 et P_5, et les points clefs P''_7 et P''_8. Les 2 cubiques pour le contour supérieur sont calculées par la méthode des moindres carrés en prenant des points du contour C_2 proches de P''_8, le point P''_8 et les coins de la bouche P_1 et P_5. Les 2 cubiques du contour inférieur sont calculées par la méthode des moindres carrés en prenant des points du contour C_3 proches de P''_7, le point P''_7 et les coins de la bouche P_1 et P_5 .

5 Résultats expérimentaux

Pour tester la performance de l'algorithme, nous utilisons la base d'images AR [13]. Elle contient des images de 126 visages (70 hommes et 56 femmes) avec différentes expressions et différentes conditions d'illumination. La largeur moyenne des bouches est de 110 pixels. La fig. 8 montre des résultats de segmentation pour cette base d'images, et ce, pour des bouches fermées et ouvertes. La dernière image est agrandie pour mieux voir la différence entre les gencives et les lèvres.
De plus, nous utilisons des séquences d'images de différentes personnes acquises dans des conditions naturelles et sans maquillage particulier. Ces images sont en RGB (8 bits/couleur/pixel) et contiennent la région du visage allant du nez au menton. La largeur moyenne des bouches est de 85 pixels. Quelques résultats pour des bouches fermées et ouvertes sont montrés sur la fig. 9.

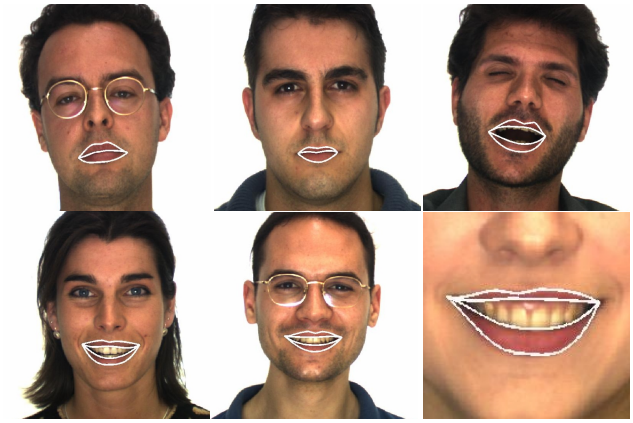

Fig. 8 - *Résultats pour des images de la base AR*

Fig. 9 - *Autres résultats*

Pour évaluer quantitativement notre algorithme dans le cas de bouches ouvertes, nous avons utilisé la méthode introduite par Wu et al. dans [14]. Nous avons annoté manuellement les contours intérieurs de 507 images de la base AR (correspondant aux conditions « smile » et « scream ») et de 94 images provenant de nos propres séquences. Si un pixel n'appartient pas à la fois à la région définie par l'annotation et à celle donnée par notre

algorithme, le pixel est considéré comme un pixel « erreur ». Le taux d'erreur est défini comme étant le rapport entre le nombre de pixels « erreur » de l'image divisé par le nombre de pixel composant la région donnée par l'annotation. La fig. 10 montre le taux d'erreur pour les 2 séries d'images. L'erreur moyenne est de 0,182 (ecart-type = 0,118) pour les images AR et de 0,188 (ecart-type = 0,068) pour les images issues des séquences.
La majorité des mauvaises détections arrive en présence de la langue, lorsque le contour n'est pas assez marqué, ou des gencives. En effet, lorsque la couleur et la texture des gencives sont trop proches de celles des lèvres, le contour est détecté entre les gencives et les dents. Quelques exemples sont montrés sur la fig. 11.

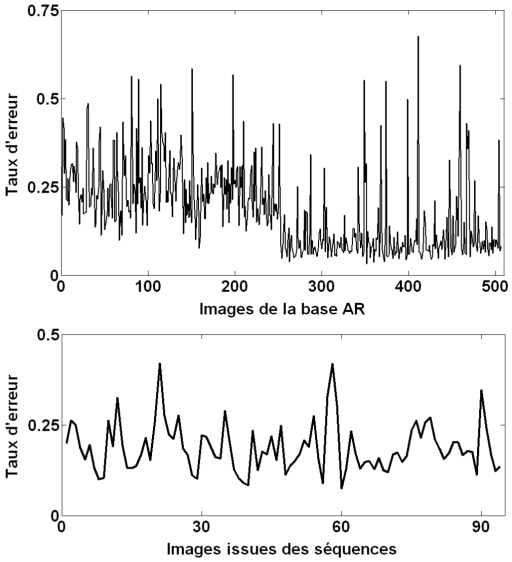

Fig. 10 – Taux d'erreur

Fig. 11 – Exemples de mauvaises segmentations

6 Conclusion

Cet article présente un algorithme de détection du contour intérieur des lèvres. La méthode consiste en la combinaison de contours actifs et de modèles paramétriques. Les contours actifs donnent les points clefs et ajustent les 2 modèles, un premier pour des bouches fermées et un second pour des bouches ouvertes. Les modèles paramétriques, composés de plusieurs cubiques, permettent d'obtenir des résultats réalistes et précis, utilisables pour des applications demandant un haut niveau de précision comme le lip reading.

Pour le moment, la décision de savoir si la bouche est fermée ou ouverte est prise manuellement et le contour intérieur est trouvé pour des images statiques. Il serait utile de détecter automatiquement si une bouche est fermée ou ouverte. En effet, pendant une conversation, la bouche alterne continuellement entre fermée et ouverte.

7 Références

[1] K. Neely, "Effect of Visual Factors on the Intelligibility of Speech", *J. Acoustical Society of America*, Vol. 28, 1956, pp. 1275-1277.
[2] W Sumby and I. Pollack, "Visual Contribution to Speech Intelligibility in Noise", *J. Acoustical Society of America*, Vol. 26, 1954, pp. 212-215.
[3] L. Zhang, "Estimation of the mouth features using deformable templates", *International Conference on Image Processing)*, Santa Barbara, CA, October 1997, pp. 328–331.
[4] B. Beaumesnil, M. Chaumont, F. Luthon, "Liptracking and MPEG4 Animation with Feedback Control", *IEEE International Conference On Acoustics, Speech, and Signal Processing*, (ICASSP'06), May 2006.
[5] T. Cootes, A. Hill, C. Taylor, J. Haslam, "Use of Active Shape Models for Locating structures in Medical Images", *Image and Vision Computing*, Vol. 12, No. 6, 1994, pp. 355-365.
[6] T. Cootes, A. Lanitis, C. Taylor, "Automatic Tracking, Coding and Reconstruction of Human Faces using Flexible Appearance Models", *IEE Electronic* Letters 30, 1994, pp.1578-1579.
[7] J. Luettin, N. Thacker, S. Beet, "Statistical Lip Modeling for Visual Speech Recognition", in *Proceedings of the 8th Eur. Signal Processing Conference* , 1996.
[8] P. Gacon, P-Y. Coulon, G. Bailly, "Non-Linear Active Model for Mouth Inner and Outer Contours Detection", *Eur. Signal Processing Conference*, Turkey, 2005.
[9] N. Eveno, A. Caplier, P-Y. Coulon, "Jumping Snakes and Parametric Model for Lip Segmentation", *International Conference on Image Processing*, Barcelona, Spain, September 2003.
[10] A. Hulbert, T. Poggio, "Synthesizing a Color Algorithm From Examples", *Science*, Vol 239, 1998, pp. 482-485.
[11] N. Eveno, A. Caplier, P-Y. Coulon, "A New Color Transformation For Lips Segmentation", *IEEE Workshop on Multimedia Signal Processing*, France, October, 2001.
[12] G. Wyszecki, W. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae, 2nd Edition*, John Wiley & Sons, Inc., New York, 1982.
[13] A.Martinez, R. Benavente, "The AR Face Database", *CVC Tech. Report* # 24, Jun. 1998.
[14] Z. Wu, P.S. Aleksic, A.K. Katsaggelos, "Lip tracking for MPEG-4 facial animation," *Proc. of IEEE International Conference on Multimodal Interfaces* (ICMI), Pittsburgh, PA, Oct. 2002.