ASSESSING THE INTRINSIC RELATIONSHIP BETWEEN FACIAL MOTION AND ACOUSTICS IN PATIENTS WITH PARKINSON'S DISEASE

Luyao Ma[1,2], Huawei Colin Li[1,3,4], Akiko Amano-Kusumoto[1,5], Willy Wong[3,4], and Pascal van Lieshout[1,2,3,6]
[1]Oral Dynamics lab, Department of Speech-Language Pathology, University of Toronto, Toronto, luyao.ma@alumni.utoronto.ca; [2]Department of Psychology, University of Toronto; [3]Institute of Biomaterials and Biomedical Engineering, University of Toronto; [4]Department of Electrical and Computer Engineering, University of Toronto; [5] OGI School of Science & Engineering, Oregon Health & Science University, Portland, USA; [6]Toronto Rehab, Toronto

1. INTRODUCTION

Gestural patterns shape the vocal tract dynamics in both visual and auditory ways, so the perception of speech is not bound to the auditory modality. From the perceivers' perspective, the coherence of observed visual information and acoustic signals is very important for comprehension[1]. Given the multi-model nature of speech perception, the congruency between facial motility and acoustic signals is an important factor in how clearly a person produces speech and how others perceive the intended message. The study described here focuses on one particular population where both facial motility and voice quality are impaired, namely individuals with Parkinson's disease.

Parkinson's disease (PD) is a common neurological condition, characterized by muscle rigidity, tremor and slowness in physical movement and is prevalent in people about 50 years of age[2]. One of the most severe consequences of PD is the lost of expressiveness in the face. The relatively weak voice in PD speakers in combination with reduced oral and facial movements makes their speech less intelligible and can have serious social consequences[3-5]. Furthermore, these adverse acoustic changes in PD which affect prosodic contrast in speech are evident in earlier stages of disease progression[3].

One way to study the congruency between speech acoustics and visual information is by mapping the relationship between acoustic data and facial motility. If the signals are highly congruent with each other, this model would provide an accurate prediction of facial motion from acoustic input. This was confirmed in a recent study from our lab using a linear multi-regression (MLR) model with data from healthy young speakers[6]. To date, it remains an open question to what extent this relation is different in people with PD.

The current study uses 3D motion data with time-aligned acoustics acquired from participants with PD, age-matched healthy speakers, and young healthy speakers. In line with previous work, we used a MLR model which has been shown to provide a good predictor model[6]. It can be hypothesized that the congruency between acoustic signals and facial motion in PD may be lower in comparison with age-matched healthy speakers and with young healthy speakers. Apart from providing important theoretical knowledge about the audio-visual relationship in speech of these populations, the results may have implications for the future development of facial motion based speech recognition software.

2. METHOD

2.1 Participants

The experimental group consisted of individuals (N=8; mean age 62.8 years) diagnosed to be in early stages of PD, recruited from the Morton & Gloria Shulman Movement Disorders Center at the Toronto Western Hospital. Two groups of healthy participants were included: an age-matched control group (OC; N=10, mean age 69.1 years) and a young control group (YC; N=10, mean age 26.3 years). Participants were excluded from the study if they shown any history of neurological disorder or disease (other than PD), orofacial musculoskeletal abnormalities, speech disorders, any history of drug and/or alcohol abuse, and hypersentivity to sunlight, as Blacklight illumination was used for motion tracking. All groups consisted of native Canadian English speakers only.

2.2 Stimuli

The speech stimuli chosen in this experiment were the same as in a previous study done on young adults and consisted of 90 sentences selected from the TIMIT and HARVARD sentence database[6]. These speech stimuli were considered to provide a representative sample of linguistic materials used in daily speech.

2.3 Procedures

Eleven glow-in-blacklight dots of face paints, about 2 mm in diameter, were applied to various locations on the face of participants. Locations include midsagittal positions on forehead, the dorsum of the nose, upper and lower lip, chin, cheeks and lip corners. The forehead position was used as a reference point for head movement correction (Figure 1). With respect to the selected gestures, F_UL represents upper lip motion relative to forehead marker; BC (bilabial closure or lip aperture) represents upper lip versus lower lip motion; F_JAW represents chin movement relative to the forehead. CHEEKS represent the motion of the left relative to the right cheeks and BP (bilabial protrusion) represent left lip corner versus right lip corner movement.

During the experiment, participants were seated in a darkened, UV illuminated room. In order to generate 3D representation, two digital camcorders were used for motion tracking located just to the left-of-center and right-of-center of the face, approximately 1 meter away and at a 45 degree angle.

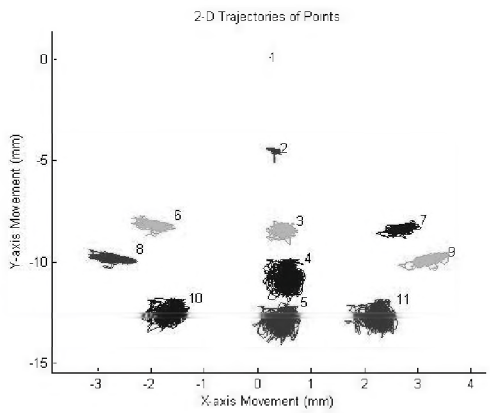

Figure 1, Example of 2-D Trajectories of the 11 markers' position for individual articulators.

The ninety sentences were randomly presented to the participants with no repetition. Participants were instructed to read sentences in a normal speaker manner. Speech acoustics was recorded with a digital voice recorder (Marantz PMD670/U1B) at 22 kHz sampling rate.

Visual and acoustic data were represented in matrix form. The acoustic information was represented by an array of 65 parameters: 16th order Linear Predictor Coefficient and 16th order Line Spectral Pairs and their first derivatives, and the root mean squared energy. In order to predict motion data from acoustic data in each time frame, an MLR analysis was performed[6]. Acoustic data corresponding to one sentence was used for testing, while remaining acoustic and visual data were used for training the MLR model. A correlation coefficient (CC) was calculated between the predicted and acquired movement[6]. This provides an index of the congruency between acoustics and facial motion.

3. RESULTS

We tested for differences in CC for GROUP (PD versus OC versus YC) and GENDER (males vs. females) for each gesture separately (F_UL, BC, F_JAW, CHEEKS, and BP) using repeated measures ANOVA with z-score transformed correlations. The original CC values for GROUP and Gesture are shown in Table 1. For BC, there was a significant GROUP effect, [$F(2,19) = 4.04$, $p = 0.03$], showing lower CC values for the PD group when compared to YC but not OC. BP also showed a GROUP effect, [$F(2,19) = 4.51$, $p = 0.02$], with YC having significantly lower correlations than OC. No other effects were found significant.

4. DISCUSSION

The findings of the current study show that individuals at an early stage of Parkinson's disease show relatively spared speech related functions, at least with the stimuli set presented in this study. However, PD subjects do show a significantly lower CC value than YC, yet not different from OC. Thus, this effect may be more of an age-related phenomenon. Lip aperture (our BC gesture) is considered the most important component in bimodal speech perception[7] and a reduced bimodal congruency may impact on speech intelligibility. This would fit with other changes in the elderly voice as reported in the literature[8].

Even though the gender effect was not significant, female PD subjects show larger differences than male PD subjects with respect to bilabial closure. Perhaps this reflects the larger fluctuations of loudness observed in female PD patients[9]. We have no clear idea what caused YC subjects to show less congruency in lip protrusion compared to older subjects, but perhaps older subjects due to the decrease in bimodal congruency in lip aperture, compensate with stronger lip movements in the horizontal dimension.

CC(STD) / Gesture	**PD**	**OC**	**YC**
F_UL	0.45 (0.16)	0.44 (0.14)	0.40 (0.09)
BC	0.54 (0.13)	0.55 (0.10)	0.66 (0.08)
F_JAW	0.52 (0.14)	0.58 (0.12)	0.64 (0.08)
CHEEKS	0.56 (0.09)	0.58 (0.07)	0.46 (0.14)
BP	0.51 (0.10)	0.57 (0.08)	0.48 (0.06)
Mean	0.51 (0.13)	0.55 (0.11)	0.53 (0.14)

Table 1, CCs of individual gestures for 3 groups

ACKNOWLEDGEMENTS

This research was undertaken, in part, thanks to funding from the Canada Research Chairs program.

REFERENCES

1. Rosenblum LD, Miller RM, Sanchez K (2007) Lip-read me now, hear me better later: cross-modal transfer of talker-familiarity effects. Psychological Science 18:392-396
2. Pinto S, Ozsancak C, Tripoliti E, Thobois S, Limousin-Dowsey P, Auzou P (2004) Treatments for dysarthria in Parkinson's disease. Lancet neurology 3:547-556
3. Cheang HS, Pell MD (2007) An acoustic investigation of Parkinsonian speech in linguistic and emotional contexts. Journal of Neurolinguistics 20:221-241
4. Miller N, Allcock L, Jones D, Noble E, Hildreth AJ, Burn DJ (2007) Prevalence and pattern of perceived intelligibility changes in Parkinson's disease. Journal of Neurology, Neurosurgery, and Psychiatry,78:1188-1190
5. Tickle-Degnen L, Lyons KD (2004) Practitioners' impressions of patients with Parkinson's disease: the social ecology of the expressive mask. Social Science & Medicine 58:603-614
6. Craig MS, van Lieshout P, Wong W (2008) A linear model of acoustic-to-facial mapping: model parameters, data set size, and generalization across speakers. Journal of the Acoustical Society of America 124:3183-3190
7. Kaynak M, Zhi Q, Cheok A, Sengupta K, Jian Z, Chung K (2004) Lip geometric features for human-computer interaction using bimodal speech recognition: Comparison and analysis. Speech Communication 43:1-2
8. Benjamin B (1997) Speech production of normally aging adults. Seminars in Speech and Language 18:135-141
9. Hertrich I, Ackermann H, Braun S, Spieker S (1996) Gender-specific vocal dysfunctions in central motor disorders. Sprache Stimme Gehor 20:169-174