Auditory-visual Perception of VCVs Produced by People with Down Syndrome: a Preliminary Study

Alexandre Hennequin, Amélie Rochet-Capellan & Marion Dohen *
Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble, France
CNRS, GIPSA-Lab, F-38000 Grenoble, France

1 Introduction

Down syndrome (DS) is the most frequent genetic disorder in humans and is present throughout society. When questioned about their child's speech, all parents of a child with DS report issues in speech intelligibility [1]. People with DS actually have better receptive than expressive speech abilities [1]. Improving speech production of people with DS is an important aspect of their quality of life. Understanding how perception of speech produced by people with DS could be improved could also have positive effects on their social integration.

Speech difficulties in people with DS originate from anatomical and physiological specificities as well as motor impairments and appear in early childhood. For example, people with DS have a smaller vocal tract and their tongue is bigger relatively to the size of their oral cavity. Other anatomical and perceptual specificities affect their ability to produce speech (see [2] for a review). All these have acoustical consequences.

To our knowledge no study has explored auditory-visual perception of speech produced by people with DS whereas it is well known that speech perception benefits from vision especially in disturbed conditions (e.g. in noise: [3]). This study aims at exploring if and how vision can improve the perception, by "ordinary" people, of speech produced by people with DS.

2 Materials and methods

AV stimuli – The stimuli were part of a larger corpus [4]. Two "ordinary" speakers (Ord; 1 female, 1 male) and two speakers with DS (1f, 1m) were selected from this corpus, all native speakers of French. The recordings took place in a sound proof room with a head mounted microphone. The task was to repeat Vowel-Consonant-Vowel (VCV) bisyllables played on a loudspeaker. Audio was acquired with a 44 100 Hz sample rate and video with a HD camera. A total of 16 stimuli were selected for each speaker with 16 different consonants covering the places and manners of articulation of French (vowel = /a/). Non-sense items were chosen to evaluate the perception of speech sounds independently from linguistic context. The mean intensity of all audio files was normalized to 70dB. A 74dB cocktail party noise (BDBRUIT database) was added to all audio files (-4dB SNR).

Perception experiment – Ten native speakers of French participated in the perceptual study. They sat in a chair in front of a 24 inch computer screen and wore headphones combined to a microphone. The experiment was divided into 3 blocks each corresponding to one modality: audio only (A), visual only (V), audiovisual (AV). The order of the blocks was randomized across participants. Each block consisted of 128 stimuli (16 VCVs * 4 speakers * 2 repetitions). The order of the stimuli was randomized within blocks and across participants. Participants were told that they would hear or see or hear and see a stimulus twice in a row and were asked to say what they had perceived.

Analyses – All the responses were manually transcribed and considered as correct if they corresponded to the intended VCV. Within-subject ANOVA was used to assess effects of modality and speaker group.

3 Results

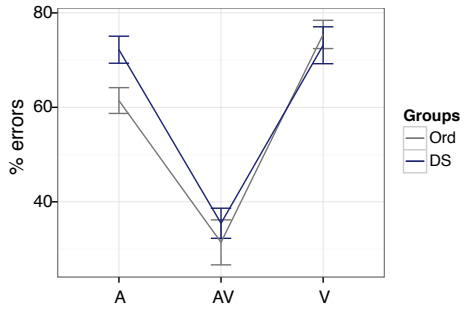

Figure 1: *Mean percentages of errors over all participants for the 3 modalities (A: audio only; V: visual only; AV: audiovisual) and the 2 groups of speakers (Ord: ordinary; DS: Down Syndrome).*

Figure 1 shows that A perception of VCVs is significantly more difficult when produced by the DS group. This is however not true for V perception. AV perception is better for both groups with, in this preliminary analysis, no reliable difference between DS and OS. This suggests that the visual information is relatively good in the two speakers with DS and that it helps compensate for differences between groups. When V is added to A, group differences disappear. In further analyses we will examine into more details confusions between consonants and compare these confusions from one group to the other. We will also extend the study to other speakers with DS. The results could be used be speech therapists to adapt readaptation procedures.

Acknowledgements

This research is part of the "Communiquons Ensemble" research project funded by the FIRAH. The authors thank the Association for Research and Social Integration of People with DS (ARIST).

References

[1] L. Kumin, "Speech intelligibility and childhood verbal apraxia in children with Down syndrome," *Down Syndrome Research and Practice*, vol. 10, no. 1, pp. 10-22, 2006.

[2] R. D. Kent and H. K. Vorperian, "Speech impairment in Down syndrome: a review," *Journal of Speech Language and Hearing Research,* vol. 56, no. 1, pp. 178-210, 2013.

[3] W. H. Sumby and I. Pollack, "Visual contribution to Speech Intelligibility in Noise," *The Journal of the Acoustical Society of America*, vol. 26, no. 2, pp. 212-215, 1954.

[4] A. Rochet-Capellan and M. Dohen M., "Acoustic characterisation of vowel production by young adults with Down syndrome," in *The International Congress of Phonetic Sciences,* Glasgow, UK, 2015 (to appear).

*e-mail: {marion.dohen,amelie.rochet-capellan}@gipsa-lab.grenoble-inp.fr