

Acoustic characterisation of vowel production by young adults with Down syndrome

Amélie Rochet-Capellan and Marion Dohen

Univ. Grenoble Alpes, GIPSA-Lab, F-38000 Grenoble, France

CNRS, GIPSA-Lab, F-38000 Grenoble, France

{amelie.rochet-capellan;marion.dohen}@gipsa-lab.grenoble-inp.fr

ABSTRACT

Down Syndrome (DS) is a frequent genetic disorder that has systematic consequences on speech articulation. The acoustic properties of speech production of people with DS have been poorly investigated in speech research. This paper reports on an acoustic analysis of vowels produced by eight native speakers of French with DS in Vowel-Consonant-Vowel (VCV) contexts. We observed more variability in duration, pitch and formants in vowels produced by people with DS compared with “ordinary” speakers. F0 was always higher for people with DS who also tended to display a larger vocalic space in VCV production than ordinary people. We interpret these results regarding motor control issues reported in previous studies involving people with DS.

Keywords: Down syndrome, Vowel space, Formants, Articulatory disorders.

1. INTRODUCTION

Down syndrome (DS), caused by the presence of an extra chromosome 21 in the genotype, is the most frequent genetic disorder in humans and is present throughout society. 15-16 out of 10 000 pregnancies lead to the birth of a child with DS [1,2]. DS induces a number of physiological disorders. Thanks to medical progress, life expectancy of people with DS has however risen from 25 years in 1983 to 49 years in 1997 [1]. It is nevertheless still a challenge to improve social integration of these people. Speech improvement is a core issue to tackle this problem.

DS is the most frequent cause of intellectual deficiency [2]. When questioned about their child’s speech, all parents of a child with DS however report speech intelligibility issues [4]. It is also well known that people with DS have better receptive than expressive speech abilities [3-4]. As intellectual abilities are very often inferred from speech abilities, speech difficulties in DS result in an overestimation of intellectual deficiency and consequently in increased difficulties in social integration. Helping

people with DS improve their speech is thus a major challenge to improve their quality of life.

Speech difficulties in people with DS originate from anatomical and physiological specificities as well as motor control impairments and appear in early childhood [3]. People with DS have a smaller vocal tract and their tongue is bigger relatively to the size of their oral cavity [6]. Other specificities such as relative size and placement of the articulators, palatal and dental abnormalities, anomalies in laryngeal structures as well as hearing loss affect the ability of people with DS to produce speech (see [7] for a review). Note that these difficulties can be improved by speech therapy, including orofacial sensori-motor stimulation starting from birth [3].

Kent and Vorperian [7] reviewed literature on speech production in people with DS over the past six decades. They considered four general areas: voice, speech sound disorders, fluency/prosody, and intelligibility. Major results can be summarized as follows. First, interest for DS in speech research has recently increased, especially concerning articulation issues. Second, results from previous studies in the four areas considered are mixed. For example, even though F0 is perceived as being lower in people with DS, results from acoustic measurements tend to show higher F0 values. Voice quality of people with DS is often described as breathy and rough but this issue is poorly quantified. People with DS make a lot of articulatory and/or phonological errors in word production. The literature also reports disfluencies as well as prosodic disturbances. Many studies, mainly based on perceptual judgements or interviews [3-5], report various problems in speech intelligibility.

Few studies have characterized the spectral properties of vowel production in people with DS. Some studies found that F1-F2 areas for different vowels overlap more in people with DS than in “ordinary” people (e.g. [8]). One or several dimensions of the vowel space would be compressed [9-11]. Results on which dimension(s) would be compressed and if such a compression actually exists are however contradictory. Some findings also suggest higher variability in formant patterns and vowel duration in people with DS [12].

As underlined in [7], “studies on vowel formant frequencies in children and adults [with DS] have been very limited and somewhat contradictory”. Most of the existing studies analysed words and not isolated syllables. Moreover, to our knowledge, no study was conducted on the acoustic analysis of vowel production of people with DS in French. This paper reports on the analysis of vowel production in VCV units in French by eight adults with DS. Based on previous studies, we expected: reduced vowel spaces for people with DS than for “ordinary” people as well as higher F0 and differences in vowel duration.

2. METHODS

2.1. Participants

Speakers were 4 females and 4 males with DS, native speakers of French, aged 19 to 34. For reference, a group of “ordinary” people was also recorded, with equivalent ages (± 5 years) and genders: 4 females and 4 males, aged 20 to 30. All participants were recorded in a sound proof room at the lab. Consent forms were read and explained to the participants with DS and to their parents. The experiment lasted less than 1h30 and all the participants received a 15 € voucher as compensation for their participation. The entire experimental procedure was approved by an ethical comity (CERNI – Approval number: 2014-03-11-41).

2.2. Set-up, material and task

Speakers sat on a chair in a sound proof room wearing a head mounted microphone (Sennheiser HSP 4). Their task was to repeat speech sounds played on a loudspeaker facing them. Auditory stimuli presentation was used because some people with DS have problems reading especially non-sense items.

One of the experimenters stayed in the room with the speaker. When the speaker had issues producing a specific VCV, the experimenter asked him/her to look at her and repeat after her. This problem never occurred with “ordinary” participants but did for all participants with DS. Some of them managed to correctly produce the target VCV after the experimenter’s intervention. Errors were however observed in several cases.

The stimuli were Vowel-Consonant-Vowel (VCV) bisyllables ($V=\{[a], [i], [u]\}$ – $C=\{[b], [d], [g], [p], [t], [k], [f], [s], [ʃ], [v], [z], [ʒ], [l], [ʁ], [m], [n]\}$). The first and second vowels in a given VCV were always identical. Vowels were chosen as the extrema of the vowel space. Consonants were chosen to cover the different places and modes of articulation of French consonants. A total of 48 different VCVs were thus

recorded for each participant. Each VCV was repeated three times. The audio prompts used for stimulation were recorded from 3 female native speakers of French aged 25, 34 and 36 in a sound proof room. The sound level of stimulation was adjusted for each participant. The recording was divided into three blocks in which the participant repeated all VCVs once. The audio prompts used in one block were all from the same speaker. The order of the VCVs was randomized within and across blocks and across participants. A total of 144 VCVs were thus recorded for each participant. The procedure was slightly different for two females with DS due to different recording periods. One participant repeated only the VCVs with a voiced consonant and was prompted by only one of the three speakers (five repetitions/VCV). The other participant followed the main procedure but was prompted with the stimuli from the three speakers randomly mixed in a single block. Due to difficulties in the recruitment of participants, we however chose to include their results in the present analyses.

Audio was acquired with a 44 100 Hz sample rate and an external sound card (Focusrite Scarlett 6i6).

2.3. Analysis and measurements

Praat [13] was used to label vowel boundaries and vocalic steady states based on the spectrogram. It was also used to extract F1 and F2 values (Burg methods, window: 25 ms, amplification: 30, time step 6 ms) as well as F0 (autocorrelation, pitch floor 75Hz, ceiling 500Hz, time step: 0.01). For formant tracking, we used the standard parameters of 5 formants in 5000Hz for males and 5 formants in 5500Hz for females. We also evaluated different parameters (4-4000, 4-5000, 6-6000), which did not improve the detection over the whole recording compared with default values. F1 and F2 were computed as the mean values 20 ms around steady state points. F0 was computed as the median value over the whole vowel.

The errors made by participants with DS depended on the participant. To get an idea of these errors, one of the authors listened to all the productions. She annotated the following errors: suppression of the first vowel ($n=2$, 1 participant), insertion of a final consonant, always after the vowel [u] ($n=13$, 1 participant), insertion of an initial consonant ($n=25$, 6 participants) mainly before or after the vowel [u] ($n=19/25$), substitution or ambiguous consonant ($n=101$, all participants, $\min=3$, $\max=28$), substitution or ambiguous first vowel (5 participants, $\min=12$, $\max=71$) or second vowel (5 participants, $\min=1$, $\max=24$). This information was not further analysed since this was not the purpose of the paper. They would require further perceptual testing.

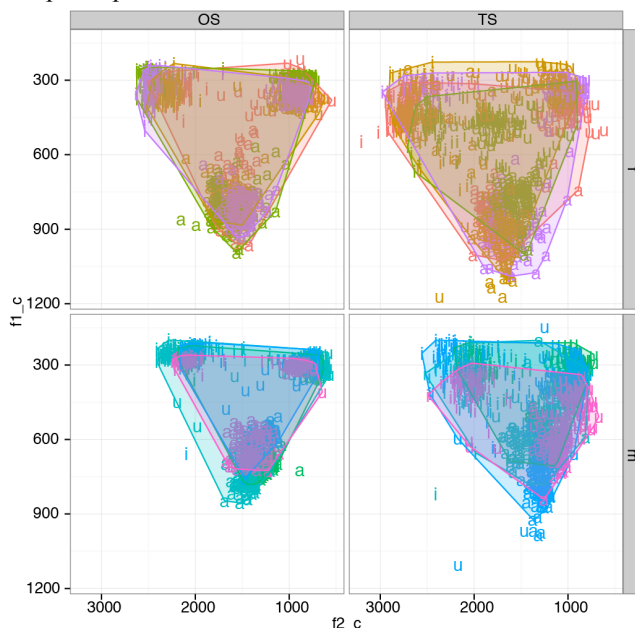
For each vowel produced we considered: its duration, F1, F2 and F0 values. For each participant, we computed the vowel space area from the convex hull defined using the function “chull” in R [14]. Statistical analyses were conducted using ANOVAs with vowel ([a] vs. [i] vs. [u]) and vowel position in the VCV (first vs. second) as within subject factors and group (TS: participant with DS vs. OS: “ordinary” participant) and gender as between subject factors. Results were considered significant for p-values below .05. Due to the small number of participants, we also considered p-values below .1 as potential tendencies for further exploration.

3. RESULTS

3.1. Formant analyses

Figure 1 provides the formant values of all vowels in the F1/F2 space for all speakers depending on group and gender. For this analysis, first and second vowels were mixed to consider pre- and post- consonant co-articulation. The area of the vowel space was computed for each speaker over all vowel occurrences. Statistical analyses show a significant effect of group on vowel space area ($F(1,12)=7$, $p=.02$), the area being larger for TS than for OS. As expected, areas were larger for females than for males ($F(1,12)=27.8$, $p<.001$).

Figure 1: Formant values of all vowels in the F1/F2 space (in Hz) for all speakers. Each data point is annotated according to the vowel supposed to be produced. OS: “ordinary” participants – TS: participants with DS – f: female – m: male.

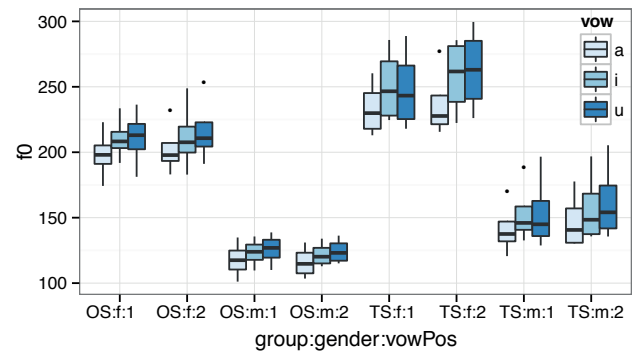


3.2. F0 analyses

Figure 2 provides F0 values of the first and second vowels for all participants depending on group and

gender. F0 is larger for TS than for OS for both females (TS: 248.4 Hz – OS: 208.4 Hz,) and males (TS: 152.6 Hz – OS: 121.5 Hz) – $F(1,12)=10.3$, $p<.008$. As expected, F0 is also greater for females (228.4 Hz) than for males (137 Hz) – $F(1,12)=68.3$, $p<.2.6e-06$ – and for the vowel [u] (188,1 Hz) than for [a] (174,2 Hz) – $F(2,24)=18.6$ $p=1.3e-05$. Vowel position has no significant effect on F0 – $F(1,12)=2.9$ $p=0.1$. Variability between subjects also appeared greater for TS than OS.

Figure 2: F0 values for the first and second vowels (in Hz). OS: “ordinary” participants – TS: participants with DS – f: female – m: male – 1st and 2nd vowel (one observation=one subject).



3.3. Duration analyses

Figure 3: Vowel durations for the first and second vowels (in s). OS: “ordinary” participants – TS: participants with DS – f: female – m: male – 1: 1st vowel – 2: 2nd vowel.

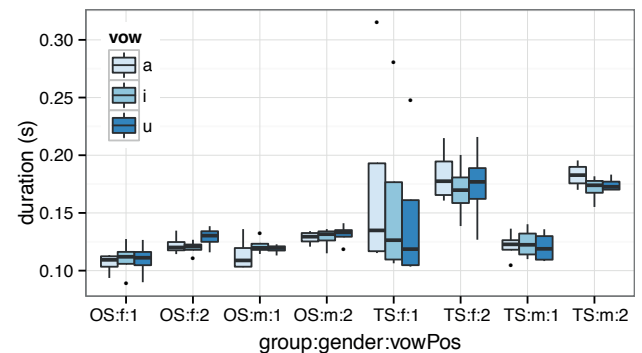


Figure 3 shows the durations of the first and second vowels for all participants depending on group and gender. The higher variability observed for females with TS is due to one participant who produced particularly long first vowels. Globally, vowels are longer for TS than OS (TS: 0,16 s – OS: 0,12 s – $F(1,12)=12.3$, $p<.004$) for both females and males. Effect of gender is not significant ($F(1,12)=0.3$, $p=.6$). For all groups and genders, the second vowel tends to be longer than the first one (1st V: 0,13 s – 2nd V: 0,15 s – $F(1,12)=4.3$, $p=.06$). There is also a tendency for [a] to be slightly longer than [i] and [u] ([a]: 0,142 s – [i]: 0,138 s – [u]: 0,138 s – $F(2,24)=2.9$, $p=.07$). Interaction between vowel and group is also

significant ($F(2,24)=11.4$, $p<.0004$): for OS Dur[a] < Dur [i] < Dur[u] and for TS Dur[u] < Dur[i] < Dur[a]. Finally, more variability in vowel durations was observed for TS than for OS.

4. DISCUSSION AND CONCLUSION

The aim of this paper was to contribute to a better characterisation of speech difficulties in people with Down syndrome (DS). As expected, greater variability in VCV production was observed for people with DS compared to “ordinary” people (OS) of comparable age. F0 was also globally higher for people with DS. By contrast with previous works, when comparing the [i]-[u]-[a] vowel space of people with DS to that of OS peers [9-10], we observed larger areas for people with DS.

The higher F0 values observed for both male and female speakers with DS compared with OS speakers is consistent with previous work [7]. It could be explained by the fact that individuals with DS are smaller and thus, may also have a smaller vocal tract. The fact that the voice of individuals with DS has been shown to be perceived lower can be related to breathing and laryngeal control. Further analyses of this phenomenon are required.

Previous studies reported a tendency towards a reduction of the vowel space in people with DS for both adults [10] and children [9]. Here the reverse pattern was observed instead: the vowel space was larger for people with DS. This difference could be related to differences in procedure and phonological material: it is possible that subjects in our study hyper-articulated more than when producing single words [10] as they were repeating VCVs produced with clear articulation. This however suggests that people with DS are able to produce vowel contrasts in hyper-articulatory conditions. This could be achieved through the production of longer vowels and thus smaller overlapping of articulatory gestures between vowels and consonant. Note that this larger space was observed even after normalization with z-scores, the phenomenon actually being increased.

Previous works also use different methods to compute the area of the vowel space, based on the identification of vowel groups. We decided not to use this method, as we were sometimes not sure that the participants with DS produced the right vowel. It is however possible that the larger space we observed is due to greater variability. Participants with DS may have explored sounds outside the phonetic space, which are not representative of their usual phonological categories. This could also be related to differences in the definition of phonemes in the perceptual space.

Previous studies involving speakers with various articulatory impairments found a relationship between size of the vowel space and speech intelligibility [15-16]. These studies were based on production of single words or full sentences. It is possible that when people with articulatory impairments produce speech under ecological conditions, their signals confirm their impairment. The advantage of using simpler non-sense phonetic material is to determine the extreme sounds people can achieve with their vocal tract. For people with intellectual disabilities, it is also an advantage not to be influenced by lexical factors since language is also affected in DS [3]. The next step is thus to determine how to include ability to control individual units in a more complex control of the speech flow.

In all the studied parameters, we observed greater variability for participants with DS than OS. This greater variability is consistent with previous work also reporting longer durations of vowels in people with DS [10]. Both increase in duration and variability could be related to specificity of muscular control in DS, which does not only concern speech but also limb control. Movements of people with DS are generally perceived to be more effortful and “hypotonic”. As pointed in [17] few studies have actually quantified hypotonia in DS. Individuals with DS display lower inertia of limb segments and probably deeper muscle relaxation at rest [17]. Applied to speech, this possible tendency for higher threshold of muscle activation and greater inertia could be related to longer durations of productions. Inertia may also explain the larger number of errors observed for the first than the second vowel in our dataset. These issues in muscular control may be even more crucial than anatomical anomalies.

Individuals with DS show great physical learning ability with appropriate training [3,17]. This points towards the necessity of providing more systematic analyses of speech acoustics and articulation in DS in order to adapt speech therapy to their specificities. Understanding articulatory abilities of individuals with DS may help improve their communicative abilities. Particular attention to adults’ abilities and their evolution over the life span should be developed.

5. ACKNOWLEDGMENTS

This research is part of the “Communications Ensemble” (Communicating Together) research project funded by the International Foundation of Applied Disability Research (FIRAH). The authors would like to thank the ARIST (Association for Research and Social Integration of People with DS) and especially the members of its board and the families. They are also grateful to the staff of the

ESAT-SAJ of the ARIST. They also thank Emilie Peyronne and Camille Pierre who participated in the recordings.

6. REFERENCES

- [1] Haute Autorité de Santé, 2007. Évaluation des stratégies de dépistage de la Trisomie 21 - Recommandation en santé publique. http://www.has-sante.fr/portail/upload/docs/application/pdf/rapport_evaluation_des_strategies_de_depistage_de_la_trisomie_2_1.pdf
- [2] Katz, G., & Lazcano-Ponce, E. (2008). Intellectual disability: definition, etiological factors, classification, diagnosis, treatment and prognosis. *salud pública de méxico*, 50, s132-s141.
- [3] Kumin, L. (2012). Early communication skills for children with Down syndrome: A guide for parents and professionals. Woodbinehouse
- [4] Kumin, L. (2006). Speech intelligibility and childhood verbal apraxia in children with Down syndrome. *Down Syndrome Research and Practice*, 10(1), 10-22.
- [5] Bunton, K., Leddy, M., & Miller, J. (2007). Phonetic intelligibility testing in adults with Down syndrome. *Down's syndrome, research and practice: the journal of the Sarah Duffen Centre/University of Portsmouth*, 12(1), 1.
- [6] Guimaraes, C.V.A., Donnelly, L.F., Shott, S.R., Amin, R.S., & Kalra, M., 2008. Relative rather than absolute macroglossia in patients with Down syndrome: implications for treatment of obstructive sleep apnea. *Pediatric Radiology*, 38, 518-523.
- [7] Kent, R.D. & Vorperian, H.K., 2013. Speech impairment in Down syndrome: a review. *Journal of Speech Language and Hearing Research* 56(1), 178-210. doi: 10.1044/1092-4388(2012/12-0148). Epub 2012.
- [8] Novak, A., 1972. The voice of children with Down's syndrome. *Folia Phoniatica* (Basel), 24, 182-194.
- [9] Moura, C.P., Cunha, L.M., Vilarinho, H., Cunha, M.J., Freitas, D., Palha, M., Puschel, S. M. & Pais-Clemente, M., (2008). Voice parameters in children with Down syndrome. *Journal of Voice*, 22, 34-42.
- [10] Bunton, K., & Leddy, M., 2010. An evaluation of articulatory working space area in vowel production of adults with Down syndrome. *Clinical Linguistics & Phonetics*, DOI: 10.3109/02699206.2010.535647.
- [11] Fourakis, M., Karlsson, H., Tilkens, C., & Shriberg, L., 2010. Acoustic correlates of nasopharyngeal resonance. In A. Botinis (Ed.) *Proceedings of the 3rd ISCA Tutorial and Research Workshop on Experimental Linguistics* (pp. 41-44). Athens: University of Athens and the International Speech Communication Association (ISCA).
- [12] Saz, O., Simon, J., Rodriguez, W.-R., Lleida, E., & Vaquero, C., 2009. Analysis of acoustic features in speakers with cognitive disorders and speech impairments. *EURASIP Journal on Advances in Signal Processing*, Article ID 159234; doi:10.1155/2009/159234.
- [13] Boersma, Paul & Weenink, David (2014). Praat: doing phonetics by computer [Computer program]. Version 5.0.42, retrieved from <http://www.praat.org/>
- [14] R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>
- [15] Audibert N. & Fougeron C. (2012) [Distorsions de l'espace vocalique : quelles mesures ? Application à la dysarthrie](#). Actes des 29èmes Journées d'Etudes sur la Parole/JEP-TALN-RECITAL 2012, pp. 217-224
- [16] Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *The Journal of the Acoustical Society of America*, 117(6), 3879-3889.
- [17] Latash, M., Wood, L., & Ulrich, D. (2008). What is currently known about hypotonia, motor skill development, and physical activity in Down syndrome.