Speech Prosody 2026
26-29 May 2026, Philadelphia, Pennsylvania, USA

How speakers living with Down syndrome produce monosyllabic interrogative and declarative sentence types – Challenges of a preliminary laboratory phonology experiment

Kornélia Juhász[1,3], Julianna Jankovics[2,3], Katalin Pirsel[1,3,4], Alexandra Markó[3]

[1]ELTE Research Centre for Linguistics,
[2]BGéSZC Öveges József Technical Secondary and Vocational School,
[3]MTA–ELTE NYTK Lendület "Momentum" Neurophonetics Research Group,
[4] ELTE Eötvös Loránd University

juhasz.kornelia@nytud.elte.hu, jankovicsjuli@gmail.com,
pirsel.katalin@nytud.elte.hu, marko.alexandra.phd@gmail.com

Abstract

This paper aims to shed light on how adult individuals living with Down syndrome (DS) produce Hungarian monosyllabic rising interrogative and falling declarative utterances. DS is one of the most common genetic disorders, however, speech characteristics of speakers living with DS are relatively under-researched, especially if prosodic patterns are concerned. Thus, monosyllabic polar interrogative and declarative utterances were recorded in the production of 8 (6 female 2 male) DS adult speakers and 8 neurotypical (female) speakers. Afterwards, DS realizations were sorted in an identification experiment based on f0 patterns with 4 evaluators. Exclusively those realizations were involved in the analysis where all 4 evaluators agreed. Interrogative and declarative f0 contours were compared by GAMMs for each speaker, accompanied by a description of minimal, maximal and mean f0. Results reveal strong individual variation in the DS speakers' production and suggest that female DS speakers show lower maximal f0 but similar mean f0 to the neurotypical group, in contrast to previous findings. In summary, most DS speakers could acoustically produce rising and falling f0 patterns similarly to controls, however, the proper elicitation of realizing the phonological contrast is still in question and requires further investigation.

Index Terms: Down syndrome, acoustic phonetics, interrogative-declarative contrast, f0

1. Introduction

The present study aims to investigate a relatively under-researched area of speech production: how adult speakers living with Down syndrome (DS speakers) contrast monosyllabic Hungarian interrogative and declarative sentence types in their production. DS is a genetic disorder resulted from a genetic error, that is, the embryo carries three copies of the chromosome 21. DS is the most common aneuploid disorder nowadays causing developmental delay and learning disabilities, including challenges of acquiring speech and linguistic functions [1, 2]. DS has a 1/800 live births chance worldwide [3], however, relative to the 25-year life expectancy in 1983, today their life span can extend to 50-60 years [2]. As there is an increased number of affected adults in the population, investigating DS speakers' speech might contribute to the visibility of this speaker group by discovering the struggles and characteristics of their prosodic production.

Studies on DS speech showed a steady increase since the 1970s, which was followed by an even more productive era in the last two decades [2]. Generally speaking, DS speakers typically show evidence of phonological disorders with varied severity within the population, however, according to [2]'s report, prosody was one of the least frequent topics for investigation. Thus, the present analysis aims to contribute to this area of research. Earlier prosodic analyses have predominantly (but not exclusively) focused on perceptual characteristics of fundamental frequency (f0), reporting that DS speakers' voice quality is nasal, hoarse, rough and breathy [4], and also displayed increased individual variability compared to neurotypical speakers [5]. As for the analysed DS groups, Moran & Gilbert [6] reported higher mean speaking f0 relative to the matching control groups. These findings were also confirmed by [4], who reported that average speaking f0 is higher for DS speakers than controls regardless of gender. Higher f0 in DS speech could occur from reduced laryngeal growth in puberty, however, cultural and psychological factors cannot be excluded shaping pitch height characteristics either.

In addition to this, DS speech is characterized by reduced efficiency of upper-respiratory function, which explains why DS speakers require twice as much effort to initiate phonation than neurotypical ones [7]. Lee et al.'s [4] results suggest limited laryngeal control at the physiological level. Consequently, DS speakers also display smaller organic f0 range, limited to 40% of those of control speakers. This tendency can be observed in the magnitude of the linguistic pitch range (f0 excursions, too, in this case: mean distance between the onset and the offset of all linguistically relevant f0 movements in the "*Rainbow passage*" text). DS speakers' f0 excursions were smaller compared to control production, signifying more monotonous intonation contours, e.g., shallower declination lines [4: 86]. It must be noted, however, that the upper-mentioned analysis [4] provided results on declarative realizations exclusively, and did not investigate how different sentence types are contrasted via f0 differences.

In this regard, even though the present analysis focuses on adults, [8]'s results yet could serve as a basis for the present study: DS children encountered problems with elevating f0 in interrogative utterance-final position. The authors concluded that the problematic nature of the utterance-final rise could be attributed to the absence/impairment of higher-level pragmatic competence (required in their experiment) and also the fact that falling contours are predominant over rising profiles. It is worth mentioning, moreover, that DS children's prosodic perception is significantly better than production, e.g., in [1]'s study, identification of rise-fall and fall-rise intonation patterns did not pose problems, whereas the production of these f0 curves was challenging. Lee [4] even suggests that in the case of adult DS speakers, prosody is discreet from other linguistic abilities, that is, there is no significant correlation between expressive-perceptive linguistic abilities and prosodic skills. Even though

10.21437/SpeechProsody.2026-179

the main focus here is on the prosodic production of speakers living with DS, as the presented results were obtained in a laboratory phonology experiment, it is also essential to briefly summarize the common neuropsychological disorders that could hinder and complicate recording sessions. Among DS speakers (children or adult), according to [9: 2]'s summary, it is common to experience repetitive behaviors [10], poor short-term memory [11], attention deficits [12], and information integration difficulties [13], leading to challenges in elicitation of prosodic patterns.

In summary, based on previous results we hypothesized that producing diverse prosodic patterns, and in this manner differentiation between illocutionary acts in production poses problems to adult DS speakers. In the case of a rising-falling contour these problems might be enlarged by the fact that rising f0 patterns are relatively rare in everyday communication. This analysis hereby presents the first steps of a wider investigation on prosodic contrast realizations among DS speakers. The study does not intend to incorporate a direct psycholinguistic approach explaining the underlying linguistic processes of DS speakers, rather provides an exploratory and preliminary research, focusing on the static and dynamic acoustic characteristics of the f0 curves contrasting Hungarian monosyllabic polar interrogative f0 pattern (characterized by a rising phase) and declarative f0 pattern (displaying a descending curve) [14, 15]. The study also presents methodological challenges and problems encountered in the attempt of elicitation of the upper-mentioned utterance types among DS speakers.

2. Method

The production of 8 adult DS speakers (six female: DS00*X* = 1, 5, 6, 7; two male: DS00*X* = 3, 8; aged between 23 and 40 years) was recorded in line with eight control speakers (all female, KO0*XX* = 02, 03, 04, 05, 06, 07, 08, 10, aged between 19 and 23 years). The majority of the DS speakers attended vocational school and are working and/or attending educational theater-related activities. One DS speaker wore a hearing aid, others did not report any hearing impairment.

Three CVC structured monosyllabic utterances were recorded in syntactically unmarked polar interrogative and declarative sentence types (see Table 1), with 3 repetitions in random order (18 realizations/speaker). The participants were asked to read aloud the utterance displayed in the monitor (see "Stimuli" in Table 1), the interpretation was aided by a corresponding picture ("Picture" in Table 1). As the intended illocutionary act was not properly realized with prosodic means by DS speakers in several cases, other elicitation methods were also applied. In one of them, target utterances were embedded into short context-providing stories, and in another one, the participants were given direct instructions after presenting the situation. Neither of the applied elicitation methods turned out to be generally successful.

All recorded realizations (without context) were evaluated and filtered in an identification experiment (carried out in Praat ExperimentMFC [16]), with 4 categories: "interrogative", "declarative", "imperative" and "problematic" (i.e., neither of above, or signaling poor quality). Those realizations were included in the acoustic analysis of which the 4 evaluators in the perception experiment (i.e., the authors of the paper) uniformly categorized. These realizations were summarized by speaker, based on the perceived illocutionary act of the utterance (Table 2). In general, producing the displayed utterance type was problematic to DS speakers in more than half of the cases because they often fell into a repetitive behavior and perseverance, repeating the same rising/falling pattern regardless of the displayed utterance, or they lost track with the instructions as a result of the relatively massive cognitive load of the task (as the recordings included other tasks, as well) and their poor short-term memory. Therefore, we decided to run the further analysis on the basis of the perceived illocutionary act, irrespective of the intention given by the original stimuli (Table 2). Finally, the realizations of DS002 were excluded, as no interrogative pattern was provided by her.

Table 2: *The number of analysed tokens per speaker (inter. = interrogative, decl. = declarative).*

Stimuli > Perceived realization	DS 001	DS 003	DS 005	DS 006	DS 007	DS 008	Total
Inter. > Inter.	6	1	7	5	3	3	**25**
Decl. > Decl.	1	6	5		3		**15**
Inter. > Decl.		5	1	1	1	2	**10**
Decl. > Inter.	3			1	1	3	**8**

The target utterances were annotated in Praat, then f0 was extracted by each 1% of the normalized duration and converted to semitones with a reference value of 50 Hz. Outlier f0 values below 15 st and above 36 st were excluded from the analysis. DS004 was further excluded from the analysis on account of the large amount of unmeasurable f0 due to irregular voicing. In this manner, we analysed 58 DS f0 contours in contrast to the 144 realizations of control speakers, displaying heavy imbalance between the two speaker groups.

Our primary aim was to deliver a dynamic analysis on f0 patterns, thus the obtained f0 curves of the two utterance types were compared by **generalized additive mixed models** (GAMMs) [17, 18, 19] for each speaker, in R [20]. In the models, f0 was analysed with respect to the normalized duration of the target monosyllable (with adaptive smooth), and the model was further complemented by a two-level parametric factor of the perceived illocutionary act, and random smooth was applied for each individual contour. In addition to this, it must be noted, that the data are quite imbalanced, with just a few realizations for DS speakers, which directly affects the quality of GAMM predictions, often resulting in wider confidence intervals of f0 curves.

Minimal, maximal, mean f0 and f0 range were extracted from the raw data for each f0 curve. Due to the small number of DS realizations, it was not possible to run inferential statistical tests, so we performed descriptive statistical analysis, i.e., we calculated the mean and standard deviation for the measured values for each speaker, statement type, and speaker group. Besides, **standard deviations (SDs)** for each static acoustic characteristics were extracted for each speaker (if applicable), and averaged along the two speaker groups and perceived illocutionary acts. In addition, as the physiological control of f0 production might be impaired among DS speakers, we also computed the **relative mean occurrence of missing values** (as an indirect and not exclusive result of irregular voicing), i.e., how many data points are missing in the by-1%-f0-measurement, divided by the number of the included utterances to relativize the percentage values (= number of missing datapoints / number of utterances involved). The presented plots were generated by ggplot2 [21].

Table 1: *Stimuli.*

Stimuli	IPA	Eng. glossa	Picture	Stimuli	IPA	Eng. glossa	Picture	Stimuli	IPA	Eng. glossa	Picture
Jár.	[ja:r]	It is working.		*Jön.*	[jøn]	It is coming.		*Mély.*	[me:j]	It is deep.	
Jár?		Is it working?		*Jön?*		Is it coming?		*Mély?*		Is it deep?	

3. Results

The **dynamic analysis** of the interrogative and declarative f0 curves revealed that – not surprisingly – the distinct f0 phases predominantly occur at the end of the target utterances in each speaker's production, when rising and falling contours are contrasted (Fig.1). Furthermore, 4 out of 8 control speakers displayed differences even in the first half of the normalized duration of the target monosyllables. This pattern applies for DS speakers, too, as 3 out of 6 speakers produced distinct f0 curve realizations in the first half of the normalized duration, disregarding the intersections or identical initial values. As it is seen on Fig.1, both DS and control speakers position their f0 curves in variable f0 registers, and several control speakers display higher overall f0 values compared to most of the DS speakers. In contrast, DS003 and DS008, i.e. the male speakers, not surprisingly, produced lower f0s (in terms of overall f0 curve position, mean, maximal and minimal f0) compared to the female speakers in the DS and the control groups (Fig. 1, 2). (In the absence of male control speakers, further comparisons were not conducted in this regard, consequently, male DS speakers were excluded from group-wise analyses, i.e., f0 range and mean SD). On the other end of the spectrum, there was one speaker for both the control (KO005) and the DS groups (DS006) displaying strikingly high maximal, minimal and **mean f0**. Interestingly, DS006's relatively high f0 mean only approximate the upper f0 level of the majority of the control speakers. The remaining three female DS speakers showed f0 means similar to the three control speakers having the lowest f0 means within the control group. In other words, DS speakers' f0 means were approximating the lower part of the range of control speakers.

There is a clear tendency in both groups that interrogatives are characterized by higher mean and maximal f0 relative to declaratives (Fig 2.), with one exception, DS005, whose declaratives showed slightly higher means relative to the interrogatives, but her f0 patterns displayed the highest variability in the DS group. If **maximal and minimal f0s are observed in a group-wise manner**, most DS speakers' maximal f0 realized, again, close to the lower end of the controls' maximums in interrogatives, as opposed to minimal f0s showing similar means in the two groups.

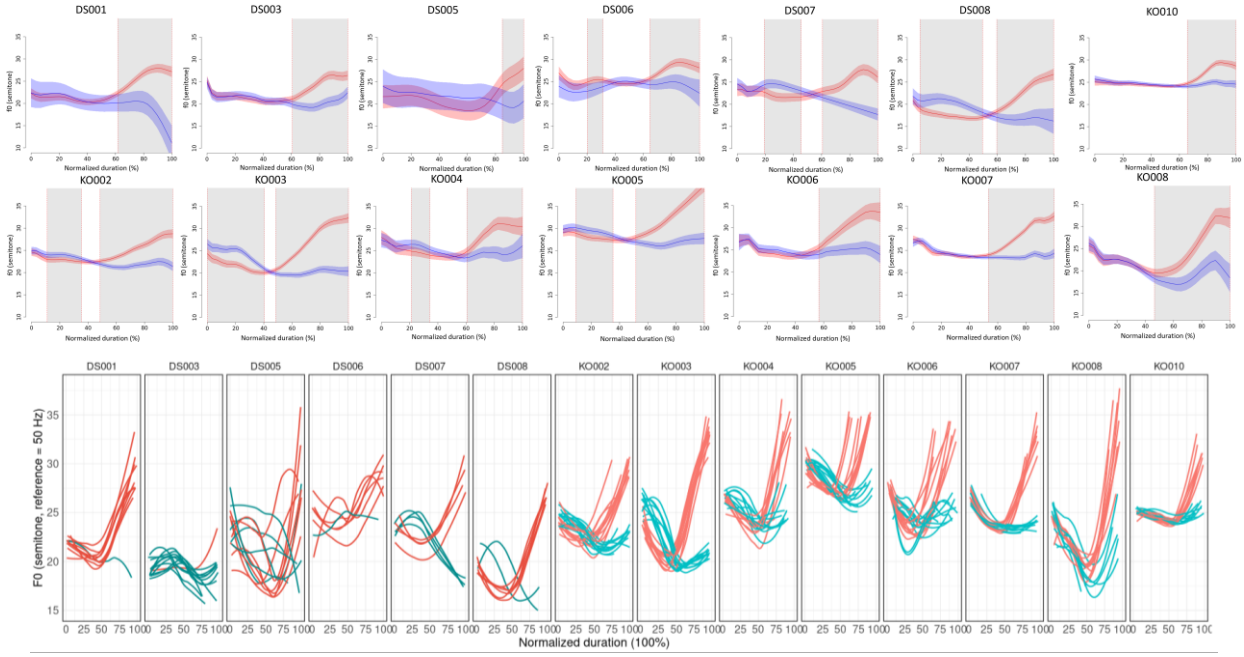

Figure 1: *Predicted f0 curves of perceived interrogative (red) and declarative (blue) utterances where grey shade signals the significant differences (above, where DS abbreviates for Down syndrome and KO for control speakers), and the individual smoothed realizations of the analysed f0 curves by each speaker (below, with same color and speaker encoding)*

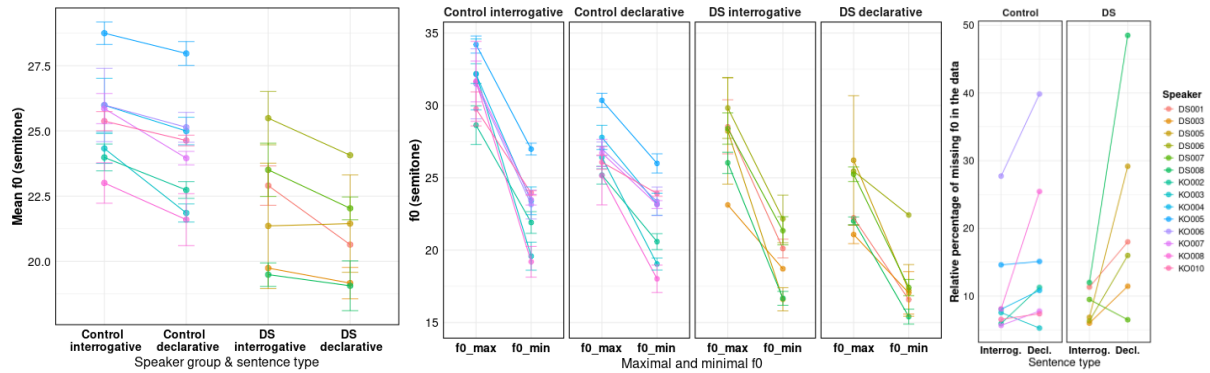

Figure 2: *Mean f0 (left), maximal and minimal f0 (middle) and relative mean occurrence of missing data (right) of interrogative and declarative utterance types by each speaker in the two speaker groups (with identical color encoding in the three plots)*

As for declaratives, DS and control speakers produced similar maximal f0 means (with an exception of DS003 realizing extremely low male-like values), as opposed to the minimal f0 where DS speakers' means were concentrated at the bottom of the analysed f0 range (with the exception of DS006 displaying an unexpectedly narrow-ranged convex and shallow pattern (Fig. 1, 2)).

Considering the individual and smoothed realizations of each speaker on Fig. 1, the imbalance of the size of datasets between the two speaker groups is apparent, as well as the relatively large number of **missing f0 values** at many positions, especially at the end of the declarative utterances. Missing data points are expected to have an indirect and not exclusive relationship with the occurrence of irregular voicing: the right panel of Fig. 2 provides a summary on this regard. 7 out of 8 control speakers realized interrogatives with less than 15% missing f0 values. Similarly, 6 control speakers produced declaratives with less than 15% missing values, which signifies that most control speakers could produce both utterance types with relatively complete contours, despite the declarative descending phase. The ratio of missing f0 in the two utterance types in DS speakers' production was more apparent in declaratives relative to interrogatives, yet, in both speaker groups there were no speakers, whose production exceeded 50% of missing values in declaratives.

The mean f0 ranges in DS speakers (Fig. 3), contrary to previous results were wider compared to that of the control group both for interrogatives and declaratives (accompanied by higher SDs). However, in DS production, the two perceived illocutionary acts' f0 ranges fell closer to each other (diff. = 1.4 st), compared to the controls' patterns, where the difference between the means were 4.1 st. Still, the SD around the f0 ranges' means of DS speakers was high, signaling extreme variability in the realizations.

Last but not least, larger mean **SD** was observed in the case of the DS group relative to the control speakers' production, regardless of either the utterance types or the static acoustic measurement types (i.e., maximum, minimum or mean f0) were concerned (Fig. 3). In general, interrogatives yielded higher mean SDs relative to declaratives in control production, as opposed to DS speakers who displayed this pattern only in terms of the mean f0. Besides, maximal f0 showed larger variability compared to minimal and mean f0 in both speaker groups.

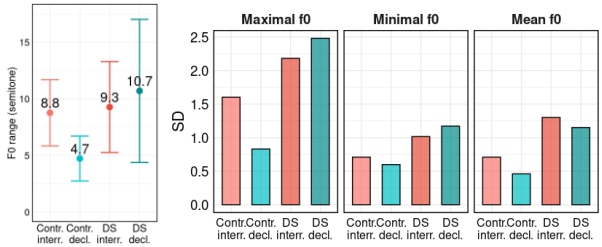

Figure 3: *Mean f0 range (left, with SDs) and mean SD of min., max. and mean f0 (right) averaged by the speakers of the two speaker groups, and along the two sentence types.*

4. Discussion

This analysis aimed to discover the acoustic characteristics of the Hungarian monosyllabic rising polar interrogative and falling declarative contrast in adult speakers' production living with Down syndrome. Originally, the analysis intended to focus on the phonological contrast, however, as we encountered many methodological challenges in the elicitation of the two sentence types, we limited our analysis on those f0 curve realizations which were perceptually identified as an interrogative or declarative illocutionary act. Thus, our results are exclusively restricted to the acoustic characteristics of identifiable rising and falling monosyllabic realizations, not particularly how the prosodic phonological contrast is taken place. As realizations were filtered on a perceptive basis, DS speakers' realizations were limited in number leading to imbalanced datasets between the DS and neurotypical control speakers' production. In this manner statistical comparisons were exclusively made in order to discover the acoustic differences of speakers' f0 curves individually, whereas static acoustic characteristics were compared in a descriptive manner. It is important to take into consideration that the individual realizations weighed more in the reduced data of DS speakers, compared to neurotypical production. In line with previous results, DS speakers' production suggests stronger intra-speaker variability compared to that of the control group. The magnitude of missing f0 values in the data (which can be indirectly associated with occurrence of irregular voicing) also stand in line with previous results suggesting limited physiological control in terms of vocal maneuvers compared to neurotypical speakers. Nevertheless, there are acoustic characteristics that are not in line with previous results: our data suggest that DS speakers' speaking f0 displays an overall similar or lower mean and maximum compared to that of the control group for both utterance types, which is different from [4, 6]'s results reporting higher f0s for DS speakers. Besides, if we take maximal and minimal values into consideration the analysis suggests control-like f0 ranges, sometimes even more extensive f0 curves relative to neurotypical speakers, which are in contrast with [4]'s study, reporting more monotonous intonation contours (e.g., shallower declination lines) [4: 86]. This might be attributed, again, to the relatively short monosyllabic utterances' nature, forcing focused, intense and short-term changes which require less prosodic planning compared to longer sentences. As rising interrogatives occur less frequently in everyday interactions relative to falling declaratives, it was expected that – in line with previous results [8] – declaratives could have been easier to produce. Yet, in our study DS speakers predominantly produced rising interrogatives. The reason behind this might lay in the methodology of the study, i.e., the monotonous reading task could easily give way to echolalia and induce repetitive behaviors with perseverance on a certain f0 contour. Besides, previous results [8] on DS children focused on polysyllabic utterances and the occurrence of utterance-final rises, which may be more susceptible to the effects of short-term memory limitations and increased cognitive load. Also, physiologically, f0 control in the higher f0 levels is easier and more sophisticated in contrary to lower f0s [22] rendering rising interrogatives easier to implement (compared to falling declaratives). Finally, the age range of DS speakers was wider, and skewed toward older ages (which might also result in lower f0 compared to younger controls) [23], but as long as simultaneous anatomical differences can also occur, the underlying factors shaping f0 production remain unexplored.

In summary, we observed extensive individual variations in DS speakers' production, which might be posed by both articulatory-physiological and neuropsychological disorders. The study also draws attention to the difficulties and limitations of the methodology of phonetic analyses of individuals living with Down syndrome, as well as contributes to the deeper understanding of the prosodic characteristics of DS speech.

5. References

[1] V. Stojanovik, "Prosodic deficits in children with Down syndrome," *J. Neurolinguistics*, vol. 24, no. 2, pp. 145–155, 2011. doi: 10.1016/j.jneuroling.2010.01.004.

[2] R. D. Kent and H. K. Vorperian, "Speech impairment in Down syndrome: a review," *J. Speech, Lang., and Hear. Res.*, vol. 56, no. 1, pp. 178–210, Feb. 2013. doi: 10.1044/1092-4388(2012/12-0148).

[3] [3] M. J. Bull, "Down syndrome," *N. Eng. J. Med.*, vol. 382, pp. 2344–2352, 2020. doi: 10.1056/NEJMra1706537.

[4] M. T. Lee, J. Thorpe, and J. Verhoeven, "Intonation and phonation in young adults with Down syndrome," *J. Voice*, vol. 23, no. 1, pp. 82–87, Jan. 2009. doi: 10.1016/j.jvoice.2007.04.006.

[5] G. E. Martin, J. Klusek, B. Estigarribia, and J. E. Roberts, "Language characteristics of individuals with Down syndrome," *Top. Lang. Disord.*, vol. 29, no. 2, pp. 112–132, Apr. 2009. doi: 10.1097/tld.0b013e3181a71fe1.

[6] M. J. Moran and H. R. Gilbert, "Speaking fundamental frequency characteristics of institutionalized adults with Down's syndrome." *Am. J. Ment. Defic.*, vol. 83, pp. 248–252, 1978.

[7] M. Pryce, "The voice of people with Down syndrome: an EMG biofeedback study," *Down Synd. Inf. Netw.*, vol. 2, no. 3, 1994.

[8] L. Zampini, M. Fasolo, M. Spinelli, P. Zanchi, C. Suttora, and N. Salerni, "Prosodic skills in children with Down syndrome and in typically developing children," *Int. J. Lang. & Comm. Disord.*, vol. 51, no. 1, pp. 74–83, Jan.–Feb. 2016.

[9] M. Corrales-Astorgano, C. González-Ferreras, D. Escudero-Mancebo, and V. Cardeñoso-Payo, "Prosodic feature analysis for automatic speech assessment and individual report generation in people with Down syndrome," *Appl. Sci.*, vol. 14, p. 293, 2024. doi: 10.3390/app14010293.

[10] E. Fucà, F. Costanzo, and S. Vicari, "Characterising repetitive behaviours in children and adolescents with Down syndrome," *J. Intellect. Disabil. Res.*, vol. 68, no. 11, pp. 1287–1300, 2024. doi: 10.1111/jir.13179.

[11] R. S. Chapman and L. Hesketh, "Language, cognition, and short-term memory in individuals with Down syndrome," *Down Syndr. Res. Pract.*, vol. 7, pp. 1–7, 2001.

[12] M. H. Martinez, X. P. Duran, and J. N. Navarro, "Attention deficit disorder with or without hyperactivity or impulsivity in children with Down's syndrome," *Int. Med. Rev. Down Syndr.*, vol. 15, pp. 18–22, 2011.

[13] R. S. Chapman, "Language development in children and adolescents with Down syndrome," *Ment. Retard. Dev. Disabil. Res. Rev.*, vol. 3, pp. 307–312, 1997.

[14] L. Varga, "The intonation of monosyllabic Hungarian yes-no questions," *Acta Linguist. Hung.*, vol. 49, no. 3–4, pp. 307–320, 2002.

[15] K. Juhász and H. Bartos, "Could L1 intonation patterns be applied in teaching Mandarin tones to atonal learners of Chinese? – An acoustic phonetic study," *Chinese as a Second Lang. Res.*, vol. 13, no. 2, pp. 157–182, 2024. doi: 10.1515/caslar-2024-2001.

[16] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer* [Computer program], Version 6.3.03, University of Amsterdam, 2022. Available: https://www.praat.org/

[17] S. N. Wood, *Generalized Additive Models: An Introduction with R*. Boca Raton: Chapman and Hall, 2017.

[18] S. N. Wood, "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models," *J. R. Stat. Soc.: Ser. B (Stat. Method.),* vol. 73, no. 1, pp. 3–36, 2011.

[19] J. van Rij, M. Wieling, R. H. Baayen, and H. van Rijn, *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs* [R package version 2.4.1], 2022.

[20] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, Version 3.6.1, 2019. Available: https://www.R-project.org/

[21] H. Wickham, W. Chang, L. Henry, T. L. Pedersen, K. Takahashi, C. Wilke, K. Woo, H. Yutani, and D. Dunnington, *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics* [R package version 3.5.1], 2024.

[22] Y. Xu, "*Contextual tonal variation in Mandarin Chinese*," Ph.D. diss., University of Connecticut, 1993.

[23] O. Tuomainen, L. Taschenberger, S. Rosen, and V. Hazan, *Speech modifications in interactive speech: effects of age, sex and noise type*, *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, vol. 377, no. 1841, 2022.