Acoustic Correlates of Nasal and Nasopharyngeal Resonance

Phonology Project Technical Report #15

Marios Fourakis

Heather Karlsson

Christie Tilkens

Lawrence Shriberg

February, 2010

Phonology Project, Waisman Center,
University of Wisconsin-Madison

Preparation of this report was supported by research grant DC00496 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health (Lawrence D. Shriberg, Principal Investigator)

Background

Phonology Project Technical Reports provide technical and substantive information on methods developed for a program of research in speech sound disorders of known and unknown origins. Primary goals of the Phonology Project are to identify etiologic origins, risk and protective factors, and diagnostic markers for eight subtypes of speech sound disorders of currently unknown origin (Shriberg, 2010).

The diagnostic instrument used in all Phonology Project studies is termed the Speech Disorders Classification System (SDCS: Shriberg et al., 2010). The SDCS includes a typologic nosology for research and practice in speech sound disorders, and an etiologic nosology for the eight putative subtypes of speech sound disorders of currently unknown origin. Data reduction methods include both perceptual and acoustic methods. Perceptual methods for narrow phonetic transcription of speech are based on extensions to the system described in *Clinical Phonetics* (Shriberg & Kent, 2003). Perceptual methods to code speakers' prosody and voice are based on extensions to the system described in The Prosody-Voice Screening Profiles (PVSP: Shriberg, Kwiatkowski, & Rasmussen, 1990) and Phonology Project Technical Report No. 1 (Shriberg, Kwiatkowski, Rasmussen, Lof, & Miller, 1992). Recent methodological focus of the Phonology Project has been on identifying acoustic correlates for all segmental and suprasegmental variables currently transcribed or prosody-voice coded using perceptual methods. The References section includes citations for these published and unpublished papers.

The present report describes acoustic correlates for two PVSP codes used to identify inappropriate resonance: PVSP Code 30: Nasal and PVSP Code 32: Nasopharyngeal. The PVSP manual includes extended discussions of each code and provides perceptual guidelines and audio exemplars to train listeners to identify each code. An acoustic correlate for the third resonance code, PVSP Code 31: Denasal, was not sought because its strong association with transient upper-respiratory congestion attenuates its diagnostic accuracy for subtypes of speech sound disorders. As cited previously, additional acoustic and psychometric information on resonance and all other PVSP codes is provided in Technical Report No. 1 of this series of laboratory reports (Shriberg et al., 1992); examples of research with the PVSP until 2001 are summarized in McSweeny and Shriberg (2001).

An Acoustic Marker for PV30: Nasal Resonance

The PVSP manual defines nasal resonance as “inappropriate” nasality in vowel contexts, and includes audio exemplars illustrating nasalized monophthongs and diphthongs in contexts in which nasalization is inappropriate in contrast to contexts in which assimilative nasality is appropriate.

Stevens (1999) provides an extensive discussion of the effect of nasalization on vowel formants. He provides theoretical calculations of the effects of nasalization on the first (F1) and second (F2) formants of non-nasal and nasal back vowels (pp. 314-315). Two sets of values are calculated for the low back vowel /ɑ/, corresponding to two different sizes of velopharyngeal opening (and subsequent coupling of the nasal and oral tracts). The effect of the coupling is the lowering of F1 by approximately 16%, when comparing the nasal with the non-nasal vowel with 0.3 cm^2 velopharyngeal opening area, and approximately 11% with 0.8 cm^2 velopharyngeal area. The effect of the coupling on the value of F2 is a raising of F2 by 6% and 14%, respectively, for each size opening.

Method

The acoustic data reported here were obtained from speakers’ responses to the Phonology Project’s assessment protocol: the Madison Speech Assessment Protocol (MSAP; Shriberg et al., 2010). The MSAP includes a conversational speech sample (CSS) obtained in the conventional manner (McSweeny, 1998) and the Challenging Words Task (CWT: Shriberg et al., 2010). The speakers were a mother and daughter, ages 49 and 18, each of whom had a chromosome 7;13 translocation causing haploinsufficiency in the gene products of *FOXP2* (Shriberg et al., 2006). Following PVSP data reduction procedures, the conversational speech sample of each speaker had been processed to yield 24 utterances eligible for PVSP coding (i.e., these utterances were retained for analyses because they did not meet criteria for one or more of 32 exclusion codes; PVSP manual, p. 11-21). All 24 CSS utterances of both speakers were coded PV30: Nasal by a research transcriber. Additionally, the nasalized diacritic [˜] had been used in the transcripts of many imitated vowel responses to CWT stimuli.

Acoustic analyses of each of the four corner vowels occurring in the CSS and occurrences of perceptually nasalized /ɑ/ vowels in CWT words were completed. For each vowel token, the frequencies of F1 and F2 were measured from a 20 ms spectral section of the vowel centered at 50% of the vowel's duration. The spectral peaks were determined using information from a Linear Prediction Coding (LPC) spectrum, a Fast Fourier Transform (FFT) spectrum, and a formant track superimposed on the spectrogram of the utterance. The LPC used a number of coefficients appropriate to the sampling rate (SR +/- 4) for each speaker's recording. The F1 and F2 values reported by Hillenbrand, Getty, Clark, & Wheeler (1995) for 48 adult women were used as the reference values for non-nasalized vowels. The F1 and F2 measurements in Hillenbrand et al. (1995) were taken from the midpoint of the four corner vowels produced in the context /h_d/. The speakers had read three randomized lists of the words in isolation.

Results

Table 1 includes F1 and F2 means and standard deviations for the total number of eligible vowel tokens obtained from the older (Panel A) and younger (Panel B) speakers with the *FOXP2* disruption. Descriptive statistics from the 48 reference speakers' productions (Hillenbrand et al., 1995) are shown in Panel C.

Table 1. Acoustic findings for vowel tokens from two speakers with perceived nasal resonance and nasalized vowels (Shriberg et al., 2006) and from a reference group (Hillenbrand et al., 1995).

Panel A. Older speaker with *FOXP2* disruption.

		F1		F2	
Vowel	No. Tokens	*M*	*SD*	*M*	*SD*
/i/	6	402	66	2659	107
/æ/	10	777	67	1842	128
/ɑ/	10	722	73	1540	233
/u/	6	453	42	1341	154

Panel B. Younger speaker with *FOXP2* disruption.

		F1		F2	
Vowel	No. Tokens	*M*	*SD*	*M*	*SD*
/i/	2	430	0	2823	13
/æ/	3	867	43	1780	286
/ɑ/	6	725	50	1255	99
/u/	4	513	49	1422	349

Panel C. Hillenbrand et al. (1995) vowel data for adult females.

		F1		F2	
Vowel	No. Speakers	*M*	*SD*	*M*	*SD*
/i/	48	437	71	2761	147
/æ/	48	676	69	2334	159
/ɑ/	48	921	97	1525	125
/u/	48	459	39	1105	205

Figures 1 and 2 are plots of the formant data for each of the two speakers and for the reference data (Hillenbrand et al., 1995). Figure 1 includes findings for the older of the two speakers and Figure 2 includes findings for the younger speaker. The Hillenbrand mean vowel values ("**X**") are connected by a line circumscribing the vowel space for adult women. In both figures, measurements of the vowel /ɑ/ from the CSS are plotted as filled squares and measurements of /ɑ/ from the CWT are plotted as open squares. As shown in both figures, the effect of nasalization is mainly evident for F1 in /ɑ/, which is lower by approximately 200 Hz than the non-nasalized mean F1 values shown for the Hillenbrand corpus (see also Table 1). The effect of nasalization for each of the two speakers is a lowering of F1 of approximately 22% relative to non-nasalized vowels. This percentage is comparable to the 16% lower values predicted by theory in Stevens (1999) as reported above. F2 values for the older speaker (Figure 1) apparently do not differ from the reference data, but for the younger speaker (Figure 2) nasalized vowels are approximately 18% lower than the reference data for non-nasalized vowels. This latter finding is not consistent with the effects of nasalization predicted using the formula in Stevens (1999).

Figure 1. F1 and F2 measurements of the corner vowels in a Conversational Speech Sample (CSS) and in the Challenging Words Task (CWT) in the Madison Speech Assessment Protocol for the older of the two speakers. The reference vowel space for the four corner vowels is defined by the mean F1 and F2 frequencies reported in Hillenbrand et al. (1995) for adult women.

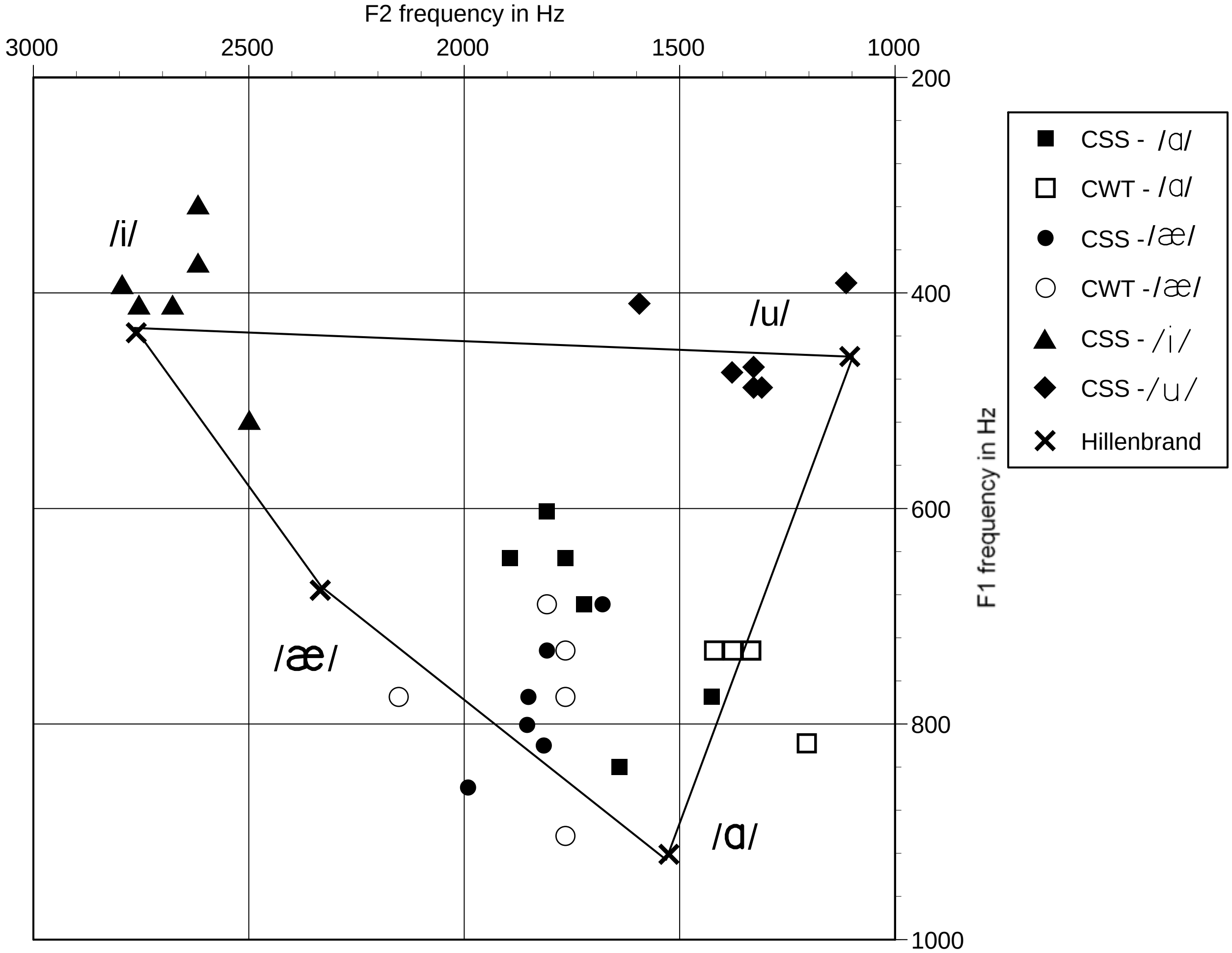
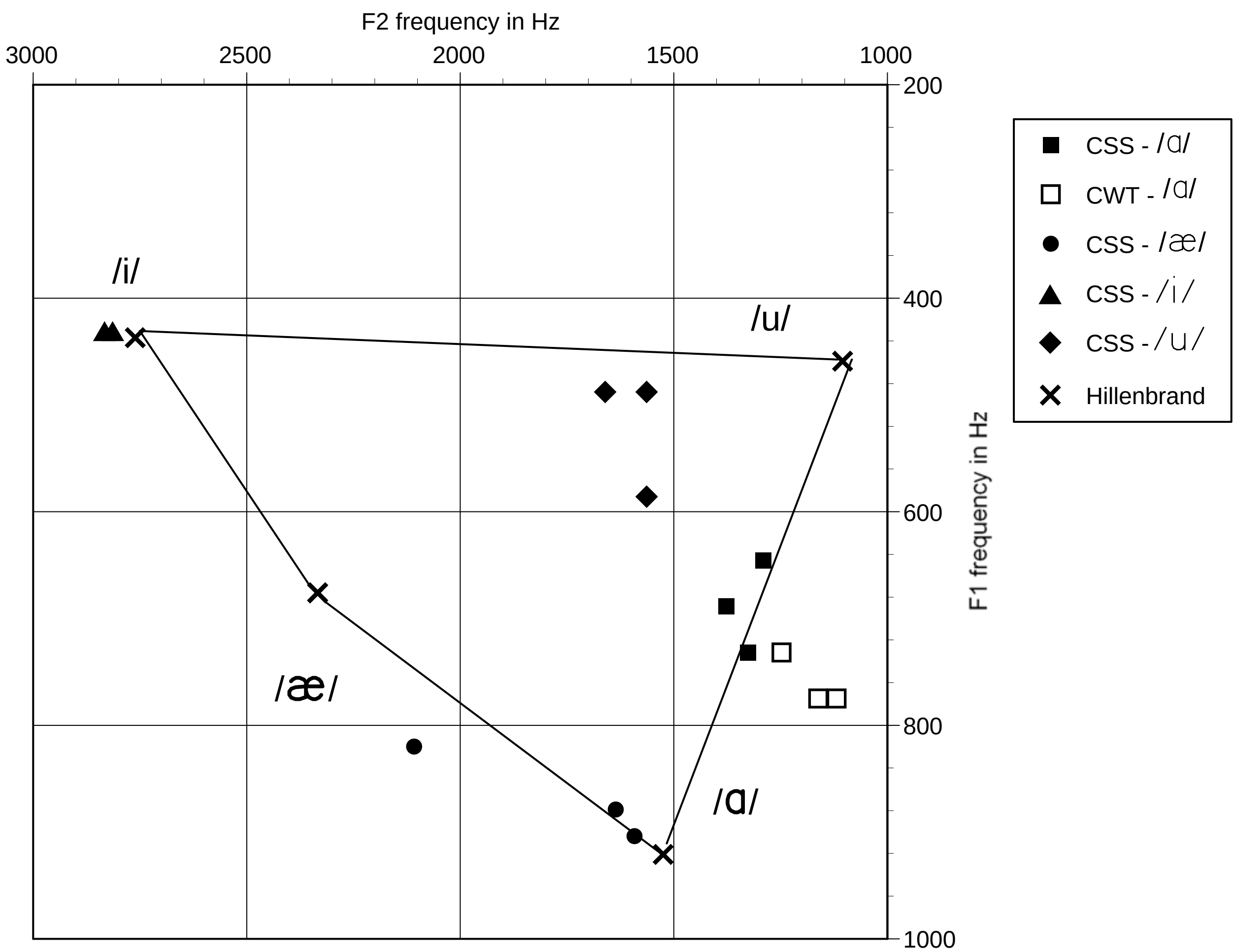


Figure 2. F1 and F2 measurements of the corner vowels in a Conversational Speech Sample (CSS) and in the Challenging Words Task (CWT) in the Madison Speech Assessment Protocol for the younger of the two speakers. The reference vowel space for the four corner vowels is defined by the mean F1 and F2 frequencies reported in Hillenbrand et al. (1995) for adult women.



Conclusion

Findings indicate that F1 lowering on /ɑ/ tokens was a robust acoustic correlate of vowel tokens perceived as nasalized from both speakers in both sampling contexts on the Madison Speech Assessment Protocol. The percentage of acoustic change from the reference data for non-nasalized vowels was acceptably close to the theoretical prediction for nasalization in Stevens (1999). These findings are interpreted as support for the use of F1 lowering on /ɑ/ vowels (i.e., values lower than one standard deviation from speakers of the same age and gender), as an acoustic correlate of perceived nasalization and nasal resonance.

An Acoustic Marker for PV32: Nasopharyngeal Resonance

The PVSP describes PV32: Nasopharygeal Resonance as a “muffled,” “back of the throat” quality consistent with the percept of “sluggish or imprecise tongue movement” sometimes used to characterize the speech of persons with Down syndrome. Several of the audio exemplars for PV32 in the PVSP manual were obtained from persons with Down syndrome. As indicated in Kent and Read’s (2002) acoustic description and elsewhere, if this “back of the throat” impression is a consequence of moving the tongue further back in the mouth than appropriate for a given vowel, the acoustic effect should be a lowering of F2.

Method

To test the sensitivity of F2 lowering as a potential acoustic correlate of perceived nasopharyngeal resonance, formant measurements of the corner vowels /i/, /æ/, /ɑ/, and /u/ were made in conversational speech samples obtained from three databases in the Phonology Project audio archives: eight 15-to 19-year-old male speakers with Fragile X syndrome, eight 15-to 17-year-old speakers with Down syndrome, and 5 typically-developing 14-year-old and 5 typically-developing 16-year-old male speakers. All tokens from each speaker came from one of their 24 CSS utterances eligible for PVSP coding (i.e., were not excluded by one or more of the 32 PVSP exclusion codes). For the speakers with Fragile X syndrome, 3-12 of the 24 utterances had been coded PV32. For the speakers with Down syndrome, at least 22 of their 24 utterances had been coded PV32: Nasopharyngeal. None of the utterances from the 10 typically-developing reference speakers had been coded PV32.

F1 and F2 frequencies for all vowels in utterances coded PV32 were measured using the same procedures as described in the first section of this paper reporting an acoustic correlate for the percept of nasal resonance. The LPC used a number of coefficients appropriate to the sampling rate (SR+/- 4) for each speaker's recording, which ranged from 20 to 44.1 kHz, depending on the speaker group and the recording date.

Results

Table 2 includes the mean and standard deviation F1 and F2 values for each corner vowel for each of the tokens obtained from speakers in each of the three groups.

Table 2. Acoustic findings for vowel tokens from two speaker groups with perceived nasopharyngeal resonance and from a reference group of typically-developing speakers.

Panel A. Conversational speech produced by adolescent speakers with Fragile X syndrome.

		F1		F2	
Vowel	No. Speakers	*M*	*SD*	*M*	*SD*
/i/	7	386	47	2233	227
/æ/	6	648	118	1786	198
/ɑ/	8	692	134	1501	114
/u/	6	396	42	1418	146

Panel B. Conversational speech produced by adolescent speakers with Down syndrome.

		F1		F2	
Vowel	No. Speakers	*M*	*SD*	*M*	*SD*
/i/	8	379	43	2234	204
/æ/	8	635	73	1803	286
/ɑ/	8	703	93	1404	120
/u/	8	405	46	1366	177

Panel C: Conversational speech produced by typically-developing adolescent speakers.

		F1		F2	
Vowel	No. Speakers	*M*	*SD*	*M*	*SD*
/i/	10	396	38	2484	204
/æ/	10	674	79	1772	122
/ɑ/	10	691	51	1305	66
/u/	10	412	43	1592	178

Figure 3 is a display of the mean F1 and F2 frequencies of the vowels produced by the speakers with Fragile X syndrome (open circles) and the typically-developing speakers (open triangles), each connected by lines to circumscribe the vowel spaces. In comparison to the values from the reference group, Fragile X speakers' lower F2 values for the high vowels /i/ and /u/ are consistent with backing of the tongue. One-tailed effect sizes for these between-group differences (i.e., to assess the directional prediction) are 1.18 (90% CI = 0.25-1.99) for /i/ and 1.05 (90% CI = 0.10-1.89) for /u/. Mean F1-F2 values for the low vowel /æ/ are generally similar for both groups. F1 and F2 values for the low vowel /ɑ/ are higher for Fragile X speakers (one-tailed effect size for F2 = 2.17; 90% CI = 1.10-3.03).

Figure 3. Mean F1 and F2 frequencies of corner vowels produced by adolescent speakers with Fragile X syndrome (circles) and typically-developing adolescent speakers (triangles).

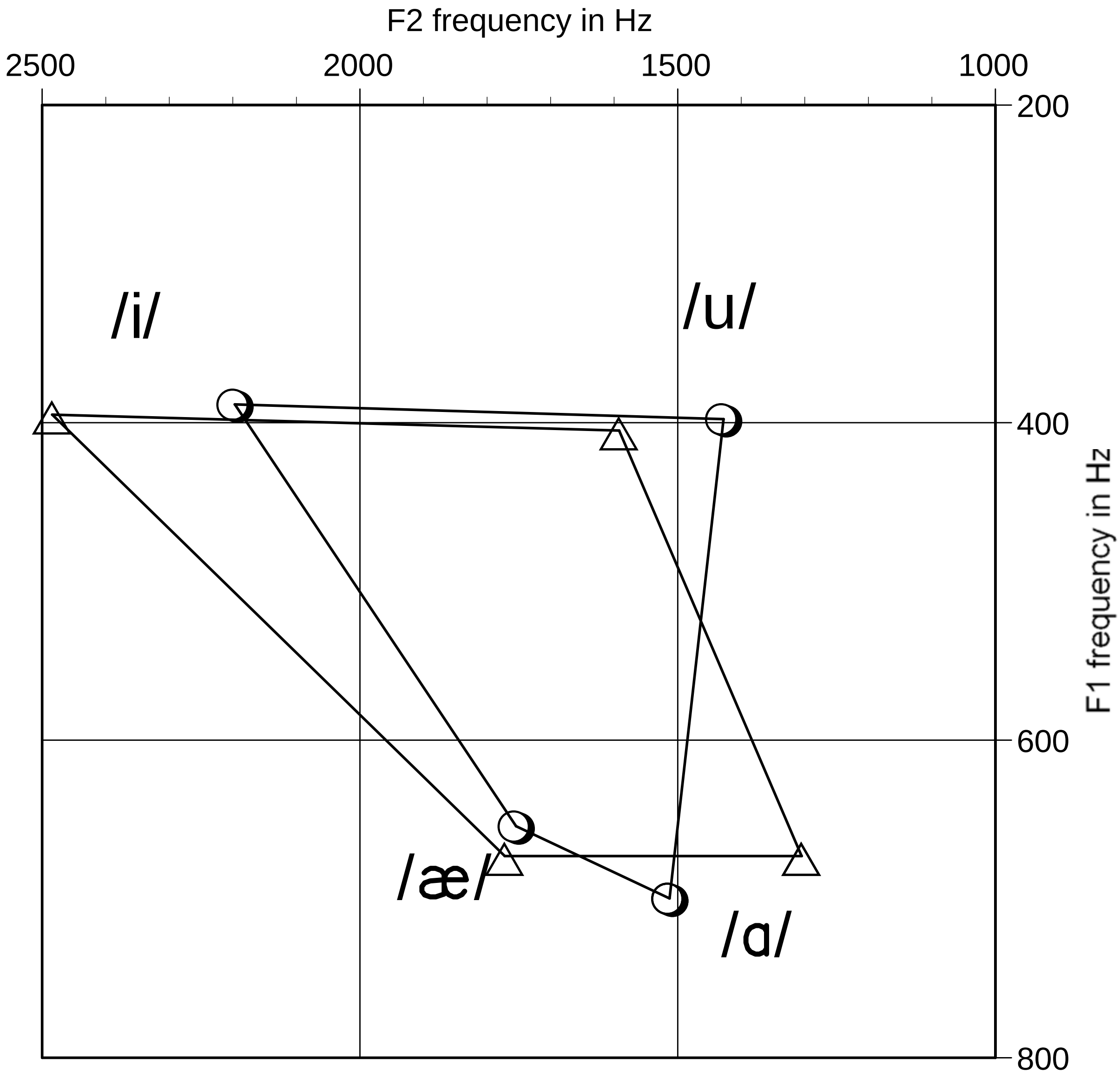
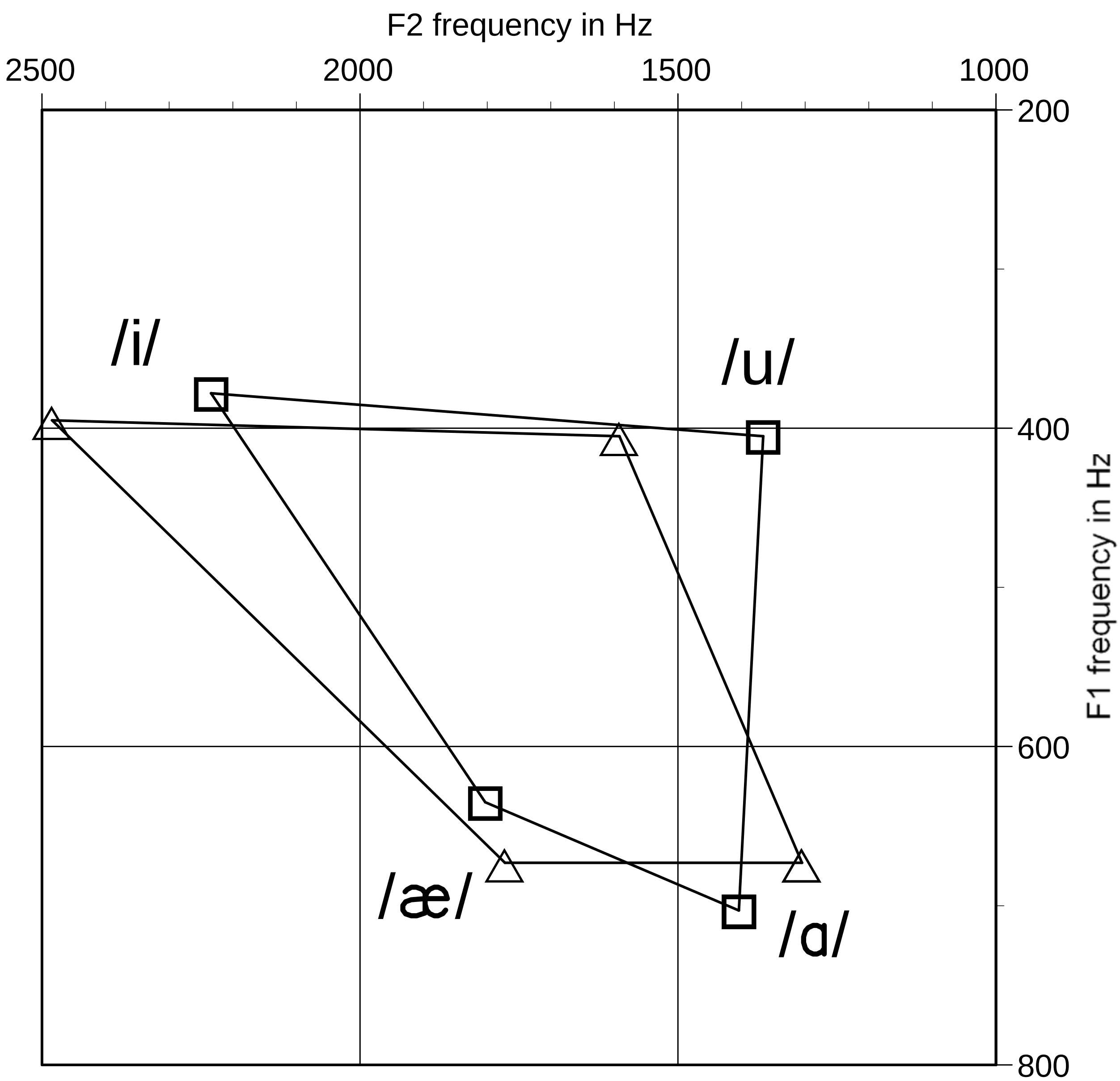


Figure 4 is a display of the means of F1 and F2 frequencies of the vowels produced by the speakers with Down syndrome (squares) and the typically-developing speakers (triangles). The between-group differences in vowel space are similar to those shown in Figure 3 for speakers with Fragile X syndrome. In comparison to values for the typical speakers, the high vowels for speakers with Down syndrome have lower F2 values, the vowel /ɑ/ has higher F1 and

F2 values, and the two speaker groups have similar values for the low vowel /æ/. The one-tailed between-group F2 effect size for /i/ is 1.23 (90% CI =0.32-2.01) and for /u/ is 1.15 (90% CI =0.26-1.93). The one-tailed between-group F2 effect size for /ɑ/ is 1.06 (90% CI =0.18-1.84).

Figure 4. Mean F1 and F2 frequencies of corner vowels produced by adolescent speakers with Down syndrome (squares) and by typically-developing adolescent speakers (triangles).



Conclusion

The acoustic findings in Table 2 and Figures 3 and 4 are interpreted as support for lowered F2 values on high vowels as a reliable acoustic correlate of the percept of nasopharyngeal resonance. In both speaker groups with complex neurodevelopmental disorders, the percept of nasopharyngeal resonance was associated with F2 values further back in the vowel space than the corresponding vowels of typically-developing adolescent speakers. Findings also supported fronting of the low-back vowel, but effects were stronger for the speakers with Fragile X syndrome than for speakers with Down syndrome. As with acoustic findings emerging for all perceptual constructs in this research series, studies in progress will attempt to cross validate these acoustic marker findings for nasal and nasopharyngeal resonance.

References and Other Phonology Project Papers on Acoustic Methods
(see also http://www.waisman.wisc.edu/phonology/)

Flipsen, P., Jr, Shriberg, L. D., Wiesmer, G., Karlsson, H., & McSweeny, J. (1998). *Acoustic reference data for /s/ in adolescents* (Tech. Rep. No. 8). Phonology Project, Waisman Center, University of Wisconsin-Madison.

Flipsen, P., Jr., Shriberg, L. D., Weismer, G., Karlsson, H. B., & McSweeny, J. L. (2000). *Acoustic data for American English /ɾ/ and /ɝ/ in typically speaking adolescents* (Tech. Rep. No. 10). Phonology Project, Waisman Center, University of Wisconsin-Madison.

Flipsen, P., Jr., Shriberg, L. D., Weismer, G., Karlsson, H. B., & & McSweeny, J. L. (2001). Acoustic phenotypes for speech-genetics studies: Reference data for residual /ɝ/ distortions. *Clinical Linguistics and Phonetics*, *15*, 603-630.

Hillenbrand, J.M., Getty, L.A., Clark, M.J., &Wheeler, K. (1995). Acoustic characteristics of American vowels. *Journal of the Acoustical Society of America*, *97*, 3099-3111.

Karlsson, H. B., Shriberg, L. D., Flipsen, P., Jr., & McSweeny, J. L. (2002). Acoustic phenotypes for speech-genetics studies: Toward an acoustic marker for residual /s/ distortions. *Clinical Linguistics and Phonetics*, *16*, 403-424.

Karlsson, H. B., Shriberg, L. D., Scheer, A., & Nadler, C. J. (2003). *Acoustic moments data for palatalized and dentalized sibilant productions from speech delayed children with and without histories of otitis media with effusion* (Tech. Rep. No. 12). Phonology Project, Waisman Center, University of Wisconsin-Madison.

Kent, R. D. & Read C. (2002). *The acoustic analysis of speech*. San Diego: Singular Press.

McSweeny, J. (1998). *Procedures to obtain extended conversational speech samples for prosody-voice analysis* (Tech. Rep. No. 7). Phonology Project, Waisman Center, University of Wisconsin-Madison.

McSweeny, J. L., & Shriberg, L. D. (2001). Clinical research with the prosody-voice screening profile. *Clinical Linguistics and Phonetics*, *15*, 508-528.

Shriberg, L. (2010). *Childhood speech sound disorders: From post-behaviorism to the post-genomic era*. In R. Paul & P. Flipsen (Eds.), Speech sound disorders in children. (pp.1–34). San Diego, CA: Plural Publishing.

Shriberg, L. D., Ballard, K. J., Tomblin, J. B., Duffy, J. R., Odell, K. H., & Williams, C. A. (2006). Speech, prosody, and voice characteristics of a mother and daughter with a 7;13 translocation affecting FOXP2. *Journal of Speech, Language, and Hearing Research*, *49*, 500-525.

Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeny, J. L., & Nadler, C. J. (2003). A diagnostic marker for childhood apraxia of speech: The lexical stress ratio. *Clinical Linguistics and Phonetics*, *17*, 549-574.

Shriberg, L. D., Fourakis, M., Hall, S., Karlsson, H. K., Lohmeier, H. L, McSweeny, J., et al. (2010). *Extensions to the Speech Disorders Classification System (SDCS)*. Manuscript submitted for publication.

Shriberg, L. D., Flipsen, P., Jr., Karlsson, H. B., & McSweeny, J. L. (2001). Acoustic phenotypes for speech-genetics studies: An acoustic marker for residual /ɝ/ distortions. *Clinical Linguistics and Phonetics*, *15*, 631-650.

Shriberg, L. D., Green, J. R., Campbell, T. F., McSweeny, J. L., & Scheer, A. (2003). A diagnostic marker for childhood apraxia of speech: The coefficient of variation ratio. *Clinical Linguistics and Phonetics*, *17*, 575-595.

Shriberg, L. D., & Kent, R. D. (2003). *Clinical phonetics* (3rd ed.). Boston: Allyn & Bacon.

Shriberg, L. D., Kent R. D., Karlsson, H. B., McSweeny, J. L., Nadler, C. J., & Brown, R. L. (2003). A diagnostic marker for speech delay associated with otitis media with effusion: Backing of obstruents. *Clinical Linguistics and Phonetics*, *17*, 529-547.

Shriberg, L. D., Kwiatkowski, J., & Rasmussen, C. (1990). *The Prosody-Voice Screening Profile*. Tucson, AZ: Communication Skill Builders.

Shriberg, L. D., Kwiatkowski, J., Rasmussen, C., Lof, G. L., & Miller, J. F. (1992). *The Prosody-Voice Screening Profile (PVSP): Psychometric data and reference information for children* (Tech. Rep. No. 1). Phonology Project, Waisman Center, University of Wisconsin-Madison.

Stevens, K. (1999). *Acoustic phonetics*. Cambridge, MA: The MIT Press.