Poznań Studies in Contemporary Linguistics 45(2), 2009, pp. 301–311
© School of English, Adam Mickiewicz University, Poznań, Poland
doi:10.2478/v10010-009-0018-2

WHY DOES THE VOWEL SPACE AREA AS AN ACOUSTIC METRIC FAIL TO DIFFERENTIATE DYSARTHRIC FROM NORMAL VOWEL ARTICULATION AND WHAT CAN BE DONE ABOUT IT?

SHIMON SAPIR
University of Haifa
sapir@research.haifa.ac.il

MONIKA POŁCZYŃSKA
Adam Mickiewicz University, Poznań
plmonik@ifa.amu.edu.pl

YISHAI TOBIN
Ben-Gurion University of the Negev, Be'er Sheva
yishai@bgu.ac.il

ABSTRACT

Background. The vowel space area (VSA) has been used as an acoustic metric of dysarthric vowel articulation, but with varying degrees of success. Here we test the hypothesis that the failure of the VSA to differentiate dysarthric from normal vowel articulation has to do in part with statistical "noise" that is introduced by insensitivity of Euclidean distances that define the VSA to vowel centralisation.

Methods. Differences in vowel production between 5 dysarthric young men post traumatic brain injury (TBI) and 5 young men who served as healthy controls (HC) were tested with four acoustic metrics: the triangular VSA, constructed with the first (F1) and second (F2) formants of the vowels /i/, /u/, /a/, and the Euclidean distances (EDs) between the vowels /i/ and /u (EDiu), /i/ and /a/ (EDia), and /a/ and /u/ (EDau) that define the VSA. The formant frequencies of these metrics were logarithmically scaled to reduce irrelevant interspeaker variability.

Results. The VSA failed to differentiate between the TBI and HC groups, as did the the EDia, and EDau. In contrast, the EDiu effectively differentiated between groups, both statistically (unpaired t-test, $p=0.0174$) and in terms effect size (1.88, large). The significant difference was in the expected direction, indicating vowel centralisation and articulatory undershoot in the TBI speakers.

Conclusion. The VSA is likely to perform poorly as an acoustic metric of dysarthric speech because of "noise" introduced by Euclidian distances (EDs) metrics that are not sensitive to the articulatory abnormality (in this study, the EDia and EDau). Thus, rather than using the VSA, it might be more beneficial to use only those EDs that are likely to be sensitive to vowel centralisation, as might be the case with the EDiu.

KEYWORDS: Traumatic Brain Injury; speech acoustics; formant centralization; vowel space area.

1. Introduction

Acoustic analysis has been used as a means to measure, in an objective, noninvasive, quantitative, and precise manner, phonatory and articulatory functions characteristic of normal and abnormal speech. One acoustic measure, vowel space area (VSA), has been used extensively to study and measure vowel articulation impairment in dysarthric speakers. The VSA is usually calculated by the Euclidean distances (EDs) between the first (F1) and second (F2) formant coordinates of the corner vowels /i/, /u/, and /a/ (triangular VSA), or the corner vowels /i/, /u/, /a/, and /æ/ (quadrilateral VSA) in the F1–F2 plane (e.g., Blomgren et al. 1998; Liu et al. 2005). For example, in the case of a triangular VSA constructed with the vowels /i/, /u/, and /a/, the VSA is expressed mathematically as follows (adopted from Blomgren et al. 1998):

(1) $VSA = sqrt(S*(S-EDiu)(S-EDia)(S-EDau))$

where

(2) $S=(EDiu+EDia+EDau)/2$
(3) $EDiu = sqrt((F1i-F1u)^2+(F2i-F2u)^2)$
(4) $EDia = sqrt((F1i-F1a)^2+(F2i-F2a)^2)$
(5) $EDau = sqrt((F1a-F1u)^2+(F2a-F2u)^2)$

Sqrt is square root, F1i is the frequency of the first formant of the vowel /i/, F2u is the frequency of the second formant of the vowel /u/, and so on. Note that the expressions (3), (4), and (5) above are the mathematical formulas for the Euclidean distances between the vowels /i/ and /u/ (EDiu), /i/ and /a/ (EDia), and /a/ and /u/ (EDau), respectively. Note also that expressions (1) and (2) are based on expressions (3), (4) and (5), as is expression (1). Thus, the triangular VSA with the vowels /i/, /u/, and /a/ is defined by the Euclidean distances (EDs): EDiu, EDia, and EDau.

In dysarthric speech, the movements of the articulators may be reduced in force and range, resulting in “undershoot” of vowel articulation. Such undershoot is related to muscle paresis or paralysis, abnormal muscle tone, or other neuropathological mechanisms. This undershoot leads to vowel formant centralisation, i.e., formants that normally have high center frequency tend to have lower frequency, and formants that normally have low center frequency tend to have higher frequency. As the result of this centralisation, the Euclidean distances are expected to be shortened and the VSA to be compressed

(Sapir et al. 2003; Sapir et al. 2007). Conversely, with decentralisation of formants, as often occurs with effective treatment, clear speech, and hyperarticulation, the Euclidean distances (EDs) are expected to increase and the VSA to expand (Ferguson and Kewley-Port 2007; Smiljanić and Bradlow 2005). Indeed several studies have shown formant centralisation and/or compression of the VSA in dysarthric speakers relative to normal speakers (e.g. Liu et al. 2005; Sapir et al. 2003; Weismer et al. 2001). Decentralisation of vowel formants and decompression of the VSA have also been reported (e.g. Sapir et al. 2003, 2007). In some studies, a positive correlation was found between VSA and intelligibility of dysarthric speech (e.g. Higgins and Hodge 2002; Liu et al. 2005).

However, other studies have failed to find statistically significant differences between dysarthric and normal speakers on vowel-formant and/or VSA measures, although there was typically an overall trend toward vowel centralisation or VSA compression in the dysarthric speakers (e.g. Ansel and Kent 1992; Bunton and Weismer 2001; Sapir et al. 2007).

Why the VSA fails to consistently differentiate normal from abnormal vowel articulation is not clear. One possible explanation is that the VSA is highly sensitive to interspeaker variability associated with anatomical differences in size and shape of the vocal tract (Nearey 1989; Perkell and Nelson 1985; Pisoni 1981). Several studies have shown that such inter-speaker variability can be considerably reduced through speaker normalisation procedures (e.g. Nearey 1989; Strange 1989). One such normalisation procedure is the transformation of formant frequencies into a logarithmic scale. The explanations for this normalisation procedure are beyond the scope of this study and can be found elsewhere (Strange 1989). In the present study we used logarithmic scaling of formant frequencies to minimise irrelevant interspeaker variability.

Another explanation, which is the scope of the present study, is the asymmetric effects of impaired vowel articulation on vowel formants frequencies in general, and on the Euclidean distances (EDs) that define the VSA in particular. For example, Sapir et al. (2007) found that of the many vowel formant combinations studied (F1i, F2a, F2i/F1i, F1a/F1i, etc.), including the VSA, only two acoustic metrics – F2u and F2i/F2u – effectively differentiated, statistically and in terms of effect size measures, between dysarthric speakers with idiopathic Parkinson's disease (PD) and healthy controls. Weismer et al. (2001) noted that some dysarthric speakers had vowel formants that were similar to those of healthy controls, whereas other formants clearly showed vowel centralisation; in still other dysarthric speakers within the same dysarthria category, the opposite happened. Yunusova et al. (2008) reported kinematic data indicating that some dysarthric speakers with amyotrophic lateral sclerosis (ALS) had reduced range of articulatory movements (consistent with vowel centralisation), whereas other dysarthric speakers with ALS were found to have articulation movements that were larger in range than those of healthy speakers. Thus, given the high sensitivity of the VSA to interspeaker variability, including the variability associated vowel formant asymmetry, it is reasonable to expect that the VSA will fail at times to differentiate between dysarthric and healthy controls due to the statistical noise induced by such variability.

On the other hand, it is possible that across dysarthric speakers, some aspects of the VSA might show more consistent sensitivity to vowel centralisation. Specifically, several studies have shown that the most likely formant to be affected by dysarthric speech articulation is F2, and that the most likely vowel to be affected by the dysarthria is /u/ (F2u) or a combination of this vowel with the vowel /i/. For example, Sapir et al. (2007) and Moura et al. (2008) have independently noted that the most sensitive acoustic metric of articulatory vowel undershoot in the patients they studied was the F2i/F2u ratio. The F2i and F2u differ markedly from each other, with the F2i typically having high center frequency (about 2400 Hz) and F2u typically having relatively low frequency (about 1000 Hz). These differences have to do with the configuration of the vocal tract, which is related to the movements or positions of the articulators. Limitations of these articulatory movements, as often is the case with dysarthria, are likely to have a strong effect on F2i and F2u, such that the center frequency of the F2i will be markedly lower than normal and the frequency of the center frequency F2u will be markedly higher than normal. Thus, with dysarthric articulatory undershoot, the F2i/F2u is likely to be markedly reduced. One would also expect that the Euclidean distance (ED) between the vowels /i/ and /u/ (i.e., EDiu) will be markedly affected by the dysarthria for the same reasons that the F2i/F2u is markedly affected.

The purpose of the present study was to test the hypothesis that the VSA may fail to differentiate between dysarthric and normal speech because of asymmetric sensitivities of the EDiu, EDia, and EDau to vowel impairment or centralisation. We also hypothesize the EDiu will be more sensitive than the EDia and EDau to vowel centralisation for the reasons mentioned above. To test these hypotheses we elected to compare vowels produced by dysarthric individuals with post traumatic brain injury (TBI) with vowels produced by healthy speakers with normal speech (healthy controls, or HC). The acoustic data (formant frequencies measurements) of these vowels were originally obtained by one of the authors for her PhD dissertation studies (Połczyńska-Fiszer 2006).

2. Methods

2.1. Subjects

Speech data from 5 young men with mild or moderate dysarthria post TBI (mean age 26.4+5.2), and 5 healthy young men with normal speech (mean age 29.2+4.8) were analysed. The biomedical and speech/language information of the 5 patients are summarised in Table 1. The subjects were all from Poland and the speech samples were all in Polish, the native language of the participants. The TBI speakers' dysarthric speech was similar to that characteristic of post-traumatic dysarthria: single vowels are intelligible even in patients whose spontaneous speech is completely incomprehensible; poorer control of the vocal folds; no glottal stops preceding vowels in word-initial position (a common feature in Polish); shorter phrases are articulated with a greater precision than

Table 1. Biomedical and speech-language information for the 5 patients.

Patient	Gender/ Age	Length of coma	Time after awakening	Type of cerebral trauma (Computer Tomography)	Mobility	Type of post-traumatic		Tracheotomy
						dysarthria	aphasia	
P1	M 27	1.5 mo.	1 yr	Intracerebral haematoma of the right temporo-parietal lobe with a perforation of the ventrical system.	Spastic quadriparesis, unable to walk.	Mild	Moderate	Yes
P2	M 22	2 mo.	5 yr	Extensive brain oedema.	No problems.	Mild	Mild	No
P3	M 25	1.5 mo.	2 yr & 7 mo.	A trace of blood in the right parietal horn. A small hypodense region in the semiovale centre (right side).	Spastic quadriparesis, unable to walk.	Mild	Moderate	Yes
P4	M 23	3 wk.	1 yr & 4 mo.	Subdural haematoma of the right fronto-temporal lobe.	Spastic quadriparesis, unable to walk.	Moderate	Moderate	Yes
P5	M 38	1 mo. & 3 wk.	1 yr. & 5 mo.	Haemorrhagic foci in the frontal regions and in the 3rd ventricle.	Spastic quadriparesis, unable to walk.	Moderate	Severe	No

longer ones; difficulties with maximum aperture and a complete closure of the articulators; speech rate is generally slower; distorted rhythm; problems with phonation (hoarse voice and breathiness); flat intonation; and hypernasality (McHenry 2000; Połczyńska-Fiszer and Pufal 2006; Wang et al. 2005).

2.2. Speech tasks

For the acoustic study, the subjects were instructed to repeat a series of the 6 Polish vowels, three times in each series, and of these 6 vowels, the first (F1) and second (F2) formants of the vowels /i/, /u/, and /a/ were each measured and averaged. This average was then used to measure the Euclidian distances (EDs) and VSA in the present study. In addition, the subjects were recorded during conversation and while reading a paragraph. These speech tasks were used to perceptually evaluate and rank the dysarthria in the patients, and to ensure normal voice and speech in the controls.

2.3. Acoustic recording and analysis

Recordings were made in a quiet room in a hospital, with a Philips SBC MD110 microphone, situated 30 cm in front of the speaker, and connected directly with the computer. The F1 and F2 of the vowels /i/, /u/, and /a/ were measured using the PRAAT software (Boersma and Weenik 2005). The measurements were made both by inspecting the formants visually and by the automatic identification of the formants with the PRAAT software. The measurements were made by the second author. The localisation of the formants as chosen by the second author was confirmed independently by two other phoneticians.

2.4. Consent

All participants completed an IRB consent form, agreeing to take part in the study. They were cognitively competent to sign such a consent form.

2.5. Statistical analyses

Unpaired t-test (alpha set at 0.05) and pooled variance effect size (ES) were used for each of the acoustic indices (VSA, EDiu, EDia, EDau) to assess differences between the TBI and HC groups. Large effect size ($\geq$0.80, Cohen 1988) was used as a criterion for a clinically significant difference between the TBI and HC groups.

3. Results

Table 2 shows the F1 and F2 data of the vowels /i/, /u/, and /a/, as well as the means of these (standard deviation, or SD, in parentheses), of the individuals in the TBI and HC groups. Table 3 shows the VSA, EDiu, EDia, and EDau data for the individuals in the two groups, as well as the means and standard deviations, and the results of the t-test and effect size measurements of the differences between the two groups. Figure 1 shows the means and standard deviations (error bars) of the four acoustic metrics in the two groups. The asterisk in Figure 1 indicates a significant difference between the two groups. As can be seen in Figure 1 and Table 3, the VSA, EDia, and EDau do not differ significantly between the groups. In contrast, the EDiu is significantly different between the groups, with a large effect size. Note that the mean EDiu is smaller in the TBI group relative to the HC group, consistent with vowel centralisation and articulatory undershoot. The means of the VSA and EDia also have lower values in the TBI group relative to the HC, and the mean of the EDau is higher in the TBI group.

Table 2. The raw formant frequency data for each of the participants (C=controls, P=patients), and the means and standard deviations of these data.

	F1i (Hz)	F2i (Hz)	F1u (Hz)	F2u (Hz)	F1a (Hz)	F2a (Hz)
C1	361	2067	440	860	816	1329
C2	345	2202	494	843	743	1290
C3	279	2550	445	790	925	1406
C4	309	2281	488	809	892	1340
C5	243	1796	479	853	814	1205
mean=	307	2179	469	831	838	1314
SD=	48	277	25	30	72	74
P1	337	2121	234	954	746	1349
P2	324	2064	391	815	820	1337
P3	319	2240	446	1189	976	1765
P4	292	2020	358	961	833	1339
P5	203	1781	430	1090	441	1497
mean=	295	2045	372	1002	763	1457
SD=	54	169	84	143	198	185

Table 3. The EDiu, EDia, EDau, and VSA data for each of the participants, the mean, and standard deviation (SD) of these data, and the results of t-test and effect size (ES) measurements. EDiu = Euclidian distance between the vowels /i/ and /u/. EDia = Euclidian distance between the vowels /i/ and /a/. EDau = Euclidian distance between the vowels /a/ and /u/ in the F1–F2 plane. VSA = vowel space area constructed with the F1 and F2 coordinates of the vowels /i/, /u/ and /a/ in the vowel space. The formant frequencies of the metrics EDiu, EDia, EDau, and VSA were logarithmically scaled (with natural log, or Ln) before these metrics were calculated. TBI = Traumatic brain injury; HC = healthy controls.

	EDiu (LnHz)	EDia (LnHz)	EDau (LnHz)	VSA ($LnHz^2$)
C1	0.90	0.93	0.76	0.31
C2	1.03	0.94	0.59	0.27
C3	1.26	1.34	0.93	0.56
C4	1.13	1.19	0.79	0.43
C5	1.01	1.27	0.63	0.31
mean=	1.07	1.13	0.74	0.38
SD=	0.14	0.19	0.14	0.12
P1	0.88	0.91	1.21	0.40
P2	0.95	1.03	0.89	0.39
P3	0.72	1.14	0.88	0.31
P4	0.77	1.13	0.91	0.35
P5	0.90	0.80	0.32	0.13
mean=	0.84	1.00	0.84	0.32
SD=	0.10	0.15	0.32	0.11
ES=	1.88	0.77	−0.41	0.55
t-test	t_8=2.9893 p=0.0174	t_8=1.2008 p=0.2642	t_8=0.6402 p=0.5400	t_8=0.8242 p=0.4337

4. Discussion

In this study the VSA, EDiu, and EDia in the TBI group had lower means compared to the HC group, consistent with vowel centralisation in the TBI group. However, the difference between the groups was significant only for the EDiu. The EDau was higher in the TBI group relative to the HC group, which is incongruent with vowel centralisation. However, this between group difference was not significant. Thus, in this study, only

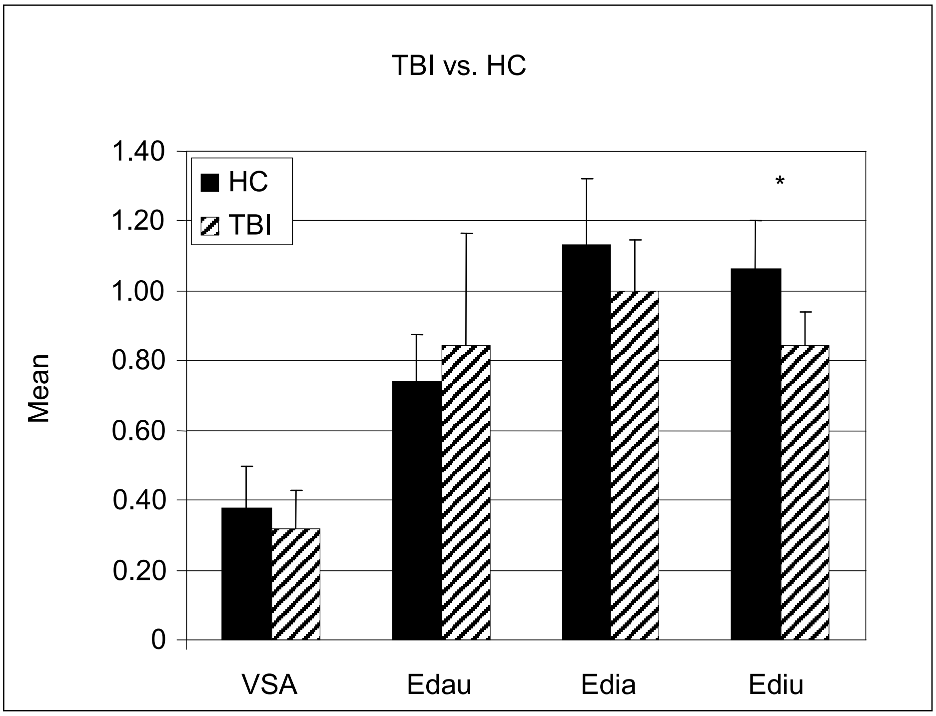

Figure 1. The means (SDs as error bars) of the metrics EDiu, EDia, EDau, and VSA in the TBI and HC groups. The asterisk indicates a significant difference ($p<0.05$) between the two groups.

the EDiu was sensitive enough to differentiate between the dysarthric and the normal vowels.

We hypothesized that the VSA may not differentiate between the TBI and HC groups because of possible statistical "noise" introduced by Euclidean distances (EDs) that are insensitive to the impairment in vowel articulation. In the present study, the EDau and EDia did not differentiate between the groups and thus most likely contributed to this "noise". It is possible that with a larger sample of participants in each of the groups, the results would have been different. For example, the mean EDia was smaller in the TBI relative to that in the HC group, but this difference was not statistically significant. It is possible that this difference between groups would be significant with a larger sample. Consequently, the VSA too may show a significant difference between the two groups.

It might be argued that the significant difference between the groups with EDiu is a spurious effect. Although this is a possibility, we believe the finding is a true reflection of the difference between the groups, and that the EDiu is simply a very sensitive metric to dysarthric vowel articulation, as we argued in the introduction. Obviously, more research is needed to confirm the present findings.

The vulnerability of the VSA to statistical noise created by some of the Euclidean distances (EDs) that define it raises questions as to the utility of this metric for the differentiation between dysarthric and normal speech. The fact that the VSA has failed in other studies to detect dysarthric vowel articulation (see review in the introduction) might be related simply to irrelevant interspeaker variability, but it might also be related to the asymmetry in the sensitivity of the vowel formants to the articulatory impairment. Thus, it seems that a more appropriate metric to differentiate dysarthric from normal vowel articulation would be the EDiu rather than the VSA. On the other hand, it is important not only to differentiate between dysarthric and normal speech, but also to characterise the nature of the vowel abnormality, and for this purpose, the profile of the different Euclidean distances and the VSA itself, as shown in Figure 1, might be more insightful.

REFERENCES

Ansel, B. and R. Kent. 1992. "Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy". *Journal of Speech and Hearing Research* 35. 296–308.

Blomgren, M., M. Robb and Y. Chen. 1998. "A note on vowel centralisation in stuttering and nonstuttering individuals". *Journal of Speech, Language, and Hearing Research* 41. 1042–1051.

Boersma, P. and D. Weenink. 2005. "Praat: Doing phonetics by computer". (Computer program, version 4.3.22.) <http://www.praat.org/>

Bunton, K. and G. Weismer. 2001. "The relationship between perception and acoustics for a high-low vowel contrast produced by speakers with dysarthria. *Journal of Speech, Language and Hearing Research* 44. 1215–1228.

Cohen, J. 1988. *Statistical power anlysis for the behavioral sciences*. (2nd ed.). Hillsdale, NJ: Earlbaum.

Ferguson, S. and D. Kewley-Port. 2007. "Talker differences in clear and conversational speech: Acoustic characteristics of vowels". *Journal of Speech, Language, and Hearing Research* 50. 1241–1255.

Higgins, C. and M. Hodge. 2002. "Vowel area and intelligibility in children with and without dysarthria". *Journal of Medical Speech-Language Pathology* 10. 271–277.

Liu, H., F. Tsao and P. Kuhl. 2005. "The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy". *Journal of the Acoustical Society of America* 117. 3879–3889.

McHenry, M. 2000. "Acoustic characteristics of voice after severe traumatic brain injury". *Laryngoscope* 100. 1156–1161.

Moura, C., L. Cunha, H. Vilarinho and M. Cunha. 2008. "Voice parameters in children with Down syndrome". *Journal of Voice* 22. 34–42.

Nearey, T. 1989. "Static, dynamic, and relational properties in vowel perception". *Journal of the Acoustical Society of America* 85. 2088–2113.

Perkell, J. and W. Nelson. 1985. "Variability in production of the vowels /i/ and /a/". *Journal of the Acoustical Society of America* 77. 1889–1895.

Pisoni, D. 1981. "Variability of vowel formant frequencies and the quantal theory of speech: A first report". *Phonetica* 37. 285–305.

Połczyńska-Fiszer M. 2006. First and second language dysarthria in TBI patients after prolonged coma. (Unpublished PhD dissertation, Adam Mickiewicz University, Poznań.)

Połczyńska-Fiszer, M. and A. Pufal. 2006. "Classification of dysarthria in Polish TBI patients using acoustic analysis". *Acta Neuropsychologia* 4. 257–285.

Sapir, S., J. Spielman, L. Ramig, S. Hinds, S. Countryman, C. Fox and B. Story. 2003. "Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on ataxic dysarthria: A case study". *American Journal of Speech Lang Pathology* 12. 387–399.

Sapir, S., J. Spielman, L. Ramig, B. Story and C. Fox. 2007. "Effects of intensive voice treatment (the Lee Silverman Voice Treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: Acoustic and perceptual findings". *Journal of Speech Language and Hearing Research* 50. 899–912.

Smiljanić, R. and A. Bradlow. 2005. "Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America* 118. 1677–1688.

Strange, W. 1989. "Evolving theories of vowel perception". *Journal of the Acoustical Society of America* 85. 2081–2087.

Wang, Y., R. Kent, J. Duffy and J. Thomas. 2005. "Dysarthria in traumatic brain injury: A breath group and intonational analysis". *Folia Phoniatrica et Logopaedica* 57. 59–89.

Weismer, G., J-Y. Jeng, J. Laures, R. Kent and J. Kent. 2001. "Acoustic and intelligibility characteristics of sentence production in neurogenic speech disorders". *Folia Phoniatrica et Logopaedica* 53. 1–18.

Yunusova, Y., G. Weismer, J. Westbury and M. Lindstrom. 2008. *Journal of Speech Language and Hearing Research* 51. 596–611.

Address correspondence to:
Shimon Sapir
Department of Communication Sciences and Disorders
Faculty of Social Welfare and Health Sciences
University of Haifa
Haifa, Mount Carmel, 39105
Israel
sapir@research.haifa.ac.il