Published in final edited form as:

J Speech Lang Hear Res. 2009 August ; 52(4): 1088–1096. doi:10.1044/1092-4388(2009/07-0167).

Comparison of acoustic and kinematic approaches to measuring utterance-level speech variability

Peter Howell, Andrew J. Anderson, Jon Bartrip, and Eleanor Bailey

Department of Psychology and Centre for Human Communications, University College London

Abstract

Purpose—The spatio-temporal index (STI) is one measure of variability. As currently implemented, kinematic data are used, requiring equipment that cannot be used with some patient groups or in scanners. An experiment is reported that addressed whether STI can be extended to an audio measure of sound pressure of the speech envelope over time, that did not need specialized equipment.

Method—STI indices of variability were obtained from lip track (L-STI) and amplitude envelope (E-STI) signals. These measures were made concurrent whilst either fluent speakers or speakers who stutter repeated “Buy Bobby a puppy” 20 times.

Results—L-STI and E-STI correlated significantly. STI reduced with age for both L-STI and E-STI. E-STI scores and L-STI scores discriminated successfully between fluent speakers and speakers who stutter.

Conclusion—The amplitude envelope over time STI scores can be used to obtain an STI score. This STI score can be used in situations where lip movement STI scores are precluded.

Keywords

Speech development; spatio-temporal index (STI); speech kinematics; stuttering; articulatory undershoot

Introduction

The spatio-temporal index or STI, (Smith, Goffman, Zelaznik, Ying & McGillem, 1995) has been used in a number of recent studies on speech-motor timing (Smith & Goffman, 2004). To calculate an STI for lower lip displacement, ten or more recordings of an utterance that involves lip closures are obtained. The individual lip-movement records are amplitude-normalized using a z transformation and time-normalized by a linear rescaling of the tokens to a common length. The standard deviation (sd) is then obtained at 2% intervals on the normalized time axis, and the computed quantities are summed to give the single STI score. The STI has been shown to be affected by the linguistic context in which the target phrase occurs, and this finding has begun to inform models of disordered (Howell, 2004; Smith & Kelly, 1997) and developmental speech control (Howell, 2004; Kleinow & Smith, 2006; Smith & Goffman, 2004). Variants on the STI measure, that involve nonlinear time scaling, have been developed by Lucero, Munhall, Gracco and Ramsay (1997), Ramsay and Silverman (1997), and Ward and Arnfield (2001).

The STI can be obtained with individual or combined movements. For example, Walsh and Smith (2002) analyzed separate upper lip, lower lip, and jaw movement trajectories in large numbers of children and adults. Smith and Zelaznik (2004) used the lip aperture and lower lip minus jaw measures to study variability in 180 children and adults. The latter study demonstrated that variability was higher for individual articulators than it was for the composite movement. Walsh, Smith and Weber-Fox (2006) also used the lip aperture variability index to study speech motor learning in children and adults.

The STI on lower lip movements has also been investigated to see whether clinical groups (principally people who stutter, PWS) differ from controls. PWS have been reported not to have different lip displacement on the “Buy Bobby a puppy” test phrase (Smith & Kleinow, 2000) although they do differ with respect to other lip movement measures (Smith & Goffman, 2004). Wohlert and Smith (2002) measured variability in labial EMG signals and computed a quantity similar to STI on lip EMG signals which they termed the EMG variability index. (EMG-VI) The EMG-VI was used to establish whether variability of this articulator changes over development (Wohlert & Smith, 2002).

Thus, lip movement alone and as part of composite movements are relevant to speech motor development and clinical conditions. In principle, however, STI can be obtained on different representations of continuous speech activity, which broadens the range of questions that can be addressed. In the case of kinematic signals, information about movement of other articulators can be processed to obtain an STI. In the current study we examined the application of STI to an audio signal (the variation in amplitude of the speech envelope (E-STI)). This is a continuous signal that represents the integrated activity of the pulmonary, laryngeal and vocal tract systems. As such, it incorporates aspects of all the major structures that must be controlled to produce speech, so if problems in control of any articulator occurs across ages or clinical groups, it should be present in this signal. We chose to examine this measure, rather than formant or pitch fundamental (F0) tracks, as points can sometimes be missed in the latter signals. Missing values can be estimated by different algorithms, which results in a continuous signal. However, in this exploratory study, we wanted to use a measure that makes as few assumptions about the signal as possible (which precluded using F0 and formant track measures with missing values corrected).

STI produces a composite measure on two dimensions which, in the case of single kinematic signals, are displacement and time. As mentioned, STI has been used on composite signals like aperture, which are not representations of the displacement of any single articulator. STI can be regarded more generally as a statistical value representing a way of quantifying variation along two dimensions, not something specific to the spatial and temporal dimensions. In the case of auditory signals like E-STI, time is equivalent to that in a kinematic signal, but the other dimension is not spatial displacement. While this makes the term STI less appropriate, we continue to use it as it has become established terminology in the literature (and it has been used previously in aperture work where it is also less appropriate).

There are several reasons for wanting to obtain STI on audio signals: 1) Audio-based signals are less expensive, less invasive and easier to record than kinematic signals. Consequently, they could be employed in clinical settings by speech-language pathologists. In this way, audio STI signals could contribute to evidence-based practice by allowing clinicians to use such signals with relatively cheap off-the-shelf equipment. 2) Movement transduction systems, which include optical tracking, electro-magnetic articulography, strain gauge and high resolution videorecording, present a range of challenges in terms of data collection and analysis that prevent widespread use in clinical environments. This problem does not apply to the same extent with audio signals, which are convenient to use. 3) It has already been

mentioned that a wide range of signals related to single oroarticulators can be extracted from the audio waveform that relate to articulation. Examples are F0 for studies in prosody or singing; frequency of the first formant to examine tongue height, and frequency of the second formant, which represents front/back position of the tongue in the vocal tract. It should be cautioned that the relationship between the audio waveform and articulatory movement is complex. 4) The degree of coordination between different articulators by different speaker groups can be examined when multiple waveforms representing simultaneous movement of different articulators are available (Max & Caruso, 1997). 5) Audio recordings can be made in scanners with suitable equipment (e.g. optical microphone systems) whereas equipment with metal components cannot be used. Thus use of measurement apparatus such as Abbs and Gilbert's (1973) head-mounted system or the Optotrack commercial system, which requires wired markers, is precluded in scanners. Sparse sampling techniques allow speech to be recorded in the quiet periods between scans so that audio signals can be obtained in these periods, which can then be used to determine how variability changes under certain conditions and to see what brain regions are associated with these changes. Signals from other movement tracking systems might also be used for scanning work such as markerless (or non-ferrous marker) video-based systems to track the lips.

The principal purpose of this work was to establish whether audio-based STIs provided measures that relate to some degree with kinematic ones. Here we specifically examined the relationship between E-STI and L-STI. This assumes variability in lip movement will affect the energy envelope, which seems reasonable given that the energy envelope represents integrated activity from all speech subsystems, including those involving articulatory movements. Although we predicted some relationship between E-STI and L-STI, it was not expected that these would correlate perfectly, as the articulatory systems operate quasi-independently. The study is preliminary to addressing issues such as: 1) potential transfer of research findings with STI to the clinical area; 2) coordination between auditory and kinematic events; and 3) central nervous system correlates of variable speech behaviour. Two independent variables known to affect L-STI were examined: 1) speakers who stutter or who are fluent (Smith & Kleinow, 2000; Smith & Goffman, 2004); and 2) speaker age (Smith & Goffman, 2004; Wohlert & Smith, 2002).

E-STI and an STI based on lower lip movement (L-STI) were obtained over time from the same utterances for people who stutter and controls (for a range of ages). The lip movement signal (L) was obtained with transducers supported by a headcage (Barlow, Cole & Abbs, 1983) and the audio waveform was used to calculate the amplitude envelope over time. The study examined: 1) Whether STIs obtained with different signals (L and E) correlated; 2) Whether STI varied with age of the speakers for each STI signal (i.e. whether spatio-temporal performance improved with age); and 3) Whether one or both STI signals distinguished fluent speakers from speakers who stutter.

Method

Participants

The participants were 25 native English speakers. The participants who stuttered were recruited as secondary referrals to a specialist stuttering clinic in the London area. They were screened at the clinic by a speech-language pathologist who confirmed stuttering. At the time of testing they were not receiving treatment (or being followed up) and had no treatment history. There were 13 speakers who stuttered, who had a mean age of 14 years 3 months (*sd* was 3 years 7 months). Two were female and the rest were male. Stuttering severity instrument version three, SSI-3, (Riley, 1994), was used to assess stuttering at the time of the test. SSI-3 ranged from very mild to severe (see Table 1 for individual data about

these speakers and the controls). There were 12 fluent speakers, who had a mean age of 16 years 11 months (*sd* was 6 years 7 months). Seven were females and the rest males. It was not possible to control for gender, but it was not expected that this factor would affect performance in this task. The two groups did not differ significantly in age by independent *t* test ($t_{23} = -1.280$, $p = 0.213$). However, there was a significant difference between the variance of the two groups (Levene's test was significant $F = 5.832$, $p = 0.024$). A non-parametric test also showed that the difference in ages of participants in the two groups was not significant (Mann-Whitney $U = 58.00$, $p = 0.276$.) This was the first time these participants had performed an experiment to obtain STI and these data have not been reported previously.

Procedure

Tests were conducted in a quiet room. Participants repeated the phrase "Buy Bobby a puppy" 20 times at their normal speech rate to ensure at least ten fluent repetitions were available. Participants were instructed to repeat the phrase as exactly as possible, paying attention to timing and amplitude control. Participants were allowed to adopt the rate and level to whatever they felt most comfortable with. Lip movement and speech recordings (used to calculate E-STI) were obtained concurrently. The microphone-to-mouth distance was monitored for each participant by the experimenter to ensure it was kept at a constant distance of approximately 20cm. Any slight differences in microphone-to-mouth distance between participants are removed when *z*-transformations are made (as required to obtain an STI). Phrases that contained word repetition, phrase repetition, pause, prolongation, part-word repetition and word break (Howell, 2007) were judged non-fluent and omitted from the analysis.

Lower lip movement data processing

Lower lip movement signals were obtained by strain gauge transducers, which measured movement along a superior-inferior plane. These were suspended from a headcage superstructure. The latter consisted of a low-mass tubular aluminum assembly. This was adjustable in order to accommodate variations and asymmetries of head size (Barlow et al., 1983). The apparatus was constructed to match that described by Abbs and Gilbert (1973). Also the headcage superstructure and transducers were positioned and stability checks were made in the same way as used by other authors (Barlow et al., 1983) and the details are not repeated here.

The cantilever/strain gauge modules were connected to an integrated circuit socket (Barlow et al., 1983). The output from each of the strain gauge transducers was passed through an amplifier and low-pass filtered (four-pole Butterworth with a cutoff of 10Hz) and captured by a PC for signal processing. An extra channel on the converter recorded voice (transduced by a Sennheiser K6 microphone). The outputs from the three strain gauges and the microphone input were low-pass filtered at 3.5kHz through a four pole Butterworth filter and sampled at 8000 samples/second, since the data acquisition card only supported a single sampling rate for all active channels; 8000 samples/second was the minimum acceptable sampling rate for the audio channel. The data were processed using MatLab version 6.5 numerical analysis software.

Audio signal data processing

The measures derived from the audio signal were obtained from Speech Filing System (SFS) software (<http://www.phon.ucl.ac.uk/resource/sfs/>). SFS has utilities that allow concurrent signals (here the audio and lip-movement signals) to be captured and displayed in alignment, one beneath the other. The onset of the first /b/ in the phrase was located from the audio track (oscillogram) and the point where voicing ceased in "puppy" was marked, again using

the audio track. These start- and end-points were used as pointers to the lip-movement track in order to select the appropriate section of this record. Other SFS utilities were used to obtain the amplitude envelope over time (E) directly from the audio record (amplitude was measured linearly). E was obtained by rectifying and low-pass filtering the signal at 15 Hz. The E track was calculated at every millisecond along the waveform as the average of the sum of the amplitude values within that millisecond frame.

STI calculation

Normalization was performed on the lip movement (the L track) and amplitude envelope (the E track) measures. Cubic spline interpolation was used to process both measures for time normalization. The length of the time axis was 100 data points for both measures. *sds* were then calculated across the normalized records. Each *sd* was calculated across the repetitions of the phrase at every other time point, giving 50 equally-spaced *sds* along the time axis overall. The STI score was then obtained as the sum of the 50 *sds*.

STI scores for different articulatory measures

Figure 1 shows the results at four steps in the analysis of L-STI (top) and E-STI (bottom) for a fluent male speaker. For each signal, the superimposed raw records are presented at top left. The panel to the right of this shows the results after filtering and amplitude normalization. The bottom left panel shows the waveforms after time normalization, and the panel to its right shows the *sd* across time after both normalization steps (STI score values for each signal are given).

Results

Correlation between STI scores for L and E

The first question addressed was whether the different STI scores correlated with one another (as would occur if a speaker's inherent variability affected all aspects of vocalization-performance). This was the case, as L-STI correlated significantly with E-STI ($r = .581$, $df = 23$, $p < .001$). The correlation on the speakers who stutter alone was significant ($r = 0.731$, $df = 11$, $p < 0.005$), and that for the fluent speakers was almost significant ($r = 0.414$, $df = 10$, $p = 0.059$). The amount of variance accounted for was between 17% and 53%. In summary, speakers who were variable on the L-STI signal also tended to be variable on the E-STI signal, and this seemed to apply to both speaker groups. The scatter plot of E-STI against L-STI is given in Figure 2.

Correlations of STI with age for L and E

The correlation of each STI signal with age was examined next. L-STI correlated negatively with age over all speakers ($r = -.774$, $df = 23$, $p < .001$), for the speakers who stutter alone ($r = -.745$, $df = 11$, $p < .005$) and for the fluent speakers alone ($r = -.811$, $df = 10$, $p < .001$). The scatter plot of L-STI ((ordinate) over age (abscissa) is shown in Figure 3. E-STI also correlated negatively over age over all speakers ($r = -.601$, $df = 23$, $p < .001$) for the speakers who stutter alone ($r = -.810$, $df = 11$, $p < .001$) and for the fluent speakers alone ($r = -.530$, $df = 10$, $p < .05$). The scatter plot of E-STI ((ordinate) over age (abscissa) is shown in Figure 4. To summarize, L- and E-STI signals decreased with age for all speakers and when speaker groups were examined separately, suggesting that spatio-temporal performance improved over this age range.

Differences across fluency groups for L and E

A two-way ANOVA was conducted with STI signal (E-STI/L-STI) as a within-groups factor, and fluency group (speakers who stutter/controls) as a between-groups factor. There

was a significant difference between fluency groups ($F_{1,23} = 6.714, p < 0.025$). The difference between STI signals and the interaction between STI signal and fluency group were not significant. Thus here, E-STI and L-STI seem equally effective at distinguishing fluency groups. This is shown in Figure 5 where the abscissa shows mean and $\pm$ one standard error for E-STI and the ordinate shows L-STI in a similar way. It can be seen that there is more overlap in the error bars for L-STI than E-STI.

Discussion

To summarize the findings, the first was that L-STI and E-STI correlated positively across participants for all participants, and for the stuttering participants but was marginal for the fluent participants. Thus, speakers who were variable on an STI score obtained on one signal tended to have high variability on the STI score obtained on the other signal. A further observation was that although L-STI and E-STI correlated significantly, the coefficients were not perfect. This indicated, as expected, that E-STI provided somewhat different information from L-STI. The second finding was that STI values (L and E) were greater for younger speakers than older ones (this applied to both speakers who stutter and controls). Finally, an ANOVA showed STI differed between fluency groups and this applied to both STI measures.

The findings suggest that audio signals can be used in future studies. There are at least three advantages in using these in selected circumstances: As E-STI was confirmed as a signal that provided similar information to kinematic signals, it could potentially be conveniently used with vulnerable groups in clinic to estimate general variability on speech control. Second, audio measures can be used when it is desirable to avoid the risk of speakers being affected by the measuring equipment. Third, use of audio signals extends the potential range of information that can be obtained about articulation.

E-STI was chosen as a test case for whether an audio signal can, in principle, be processed by STI. Thus, the purpose of this article was not specifically to promote use of E-STI. However, as this signal has proved useful in discriminating speaker characteristics known to influence lip STI data, it should be investigated further. For instance, it should be looked at under different speaking conditions, recording factors/environments, using linear versus. dB scale and so on.

The correspondence between L-STI and E-STI results with respect to the three reported findings are consistent with the views: 1) that it is appropriate to obtain STI on this audio measure; and 2) that there is some general system variability that depends on the operation of all articulatory subsystems. The general variability affects measures reflecting displacement of a single articulator and the audio signal. We begin by considering what E-STI measures and in what way this differs from L-STI, given the less than perfect correlation. As stated earlier, the E signal represents the activity of pulmonary, laryngeal and vocal tract components. The L-STI measures one of the latter signals alone. The putative origins of the different STI signals may account for the pattern seen in the correlation coefficients for the different speaker groups over age on the different STI measures. Although the correlations were significant for all speaker groups for both types of STI, L-STI correlated better with age for the fluent speakers than the speakers who stutter, and E-STI correlated better with age for speakers who stutter than for fluent speakers. The pattern with L-STI may be explained on the assumption that pulmonary and laryngeal control for fluent speakers is better established when they have reached the ages at which they were tested but less so for speakers who stutter. As there is more scope for improvement with the people who stutter, this would then lead to a better correlation with L-STI than there is for fluent speakers. For the E-STI signal, in contrast, more improvement in coordinated control

of the several subsystems making up this signal may be required for both speaker groups. The better correlation for the fluent group across ages might then indicate that coordinated control develops more systematically in this age range for the fluent speakers. This explanation of the results is consistent with the view that E-STI measures pulmonary, laryngeal and vocal tract aspects of control.

Before coordinative control can be investigated further using audio and/or kinematic measures, some additional work would be necessary. The first issue is what signals to extract in order to examine coordination between them. Kinematic signals are often difficult to obtain whereas, whilst audio signals are more straightforward to obtain, and although they have independent legitimacy (e.g. F0 and formant movements are important descriptors of speech output), they have a complicated relationship with the movement of any single articulator. The choice of what signals to use is not straightforward and depends on the particular research application: if displacement has to be measured, kinematic measurement systems should be used; for analyses like those described next, audio as well as kinematic signals can be employed.

The second issue is that the temporal and spatial components ought to be estimated separately when comparing audio and kinematic signals. This can be achieved with the nonlinear methods. Some particular questions concerning temporal control can then be addressed. For instance: Does altered feedback affect either cerebellar/basal ganglia timing mechanisms (Alm, 2004; Howell, 2002; Howell & Sackin, 2002)? Alternatively does altered feedback affect higher level linguistic monitoring processes (Levelt, 1989)? The adjustments needed to bring different signals (audio and/or kinematic ones) into alignment on the time axis could also be compared and used as an indication of inter-articulator coordination (Howell, Anderson & Lucero, in press). Examination of the residual variability after temporal variability has been extracted could be useful. We argued earlier that the non-temporal component of kinematic and audio signals can be regarded as representing aspects that are related at an abstract level even though they are not equivalent in terms of measurement units. If this is true, a testable prediction is that the spatial displacement signal from L-STI should correlate with the equivalent signal from E-STI, but not with the temporal signal from E-STI. Such a correlation might be expected in the case of lip opening and envelope amplitude, but should be investigated for other audio signals where the relationship with the L-STI or other kinematic signals is not so obvious. Once coordination measures are available and validated, they can be applied to examine hypotheses about different clinical groups. Any differences in ability to coordinate articulatory signals may indicate whether individuals who stutter have an unstable speech system that is susceptible to variability, as specified in Packman, Onslow, Richard and van Doorn's (1996) V model. Also, McHenry (2003) hypothesized that Parkinson Disease patients may be more variable in the temporal domain and that ataxic patients may be more variable in the spatial domain. Again these predictions could be tested.

As was also mentioned in the introduction, the E-STI can be measured with equipment that does not have metal components so can be used in future scanning work to establish the relationship between fluency and brain activation. The findings offer the possibility of examining motor variability in environments where it is not possible to use metal equipment (i.e. in scanners) or with participants who may not tolerate their use, such as young children. To address the former, we have made pilot recordings in an MRI scanner, which confirm that it is possible to obtain audio signals that are suitable for obtaining STI estimates.

Limitations

One limitation of the audio method is that it requires some knowledge of making recordings and use of equipment. The procedures work well in the research lab and, as observed, in

MRI scanners. However, corridor noise and any extraneous noises in a clinic setting would have to be avoided. A limitation of the study is the small number of participants in both groups. Also there is a gender imbalance between the two fluency groups which conceivably might affect results. Pair by pair age matching of control and test participants could be done rather than a post hoc check on whether ages differed between the fluency groups. For the children who stutter, variation in STI with severity level, and whether the child ultimately persists or recovers, also need to be examined.

Acknowledgments

This work was supported by grant 072639 from the Wellcome Trust to Peter Howell.

References

- Abbs JH, Gilbert B. A strain gauge transduction system for lip and jaw motion in two dimensions: Design criteria and calibration data. *Journal of Speech and Hearing Research*. 1973; 16:248–256. [PubMed: 4723981]
- Alm PA. Stuttering and the basal ganglia circuits: A critical review of possible relations. *Journal of Communication Disorders*. 2004; 37:325–369. [PubMed: 15159193]
- Barlow SM, Cole KJ, Abbs JH. A new head-mounted lip-jaw movement transduction system for the study of motor speech disorders. *Journal of Speech and Hearing Research*. 1983; 26:283–288. [PubMed: 6887815]
- Howell, P. The EXPLAN theory of fluency control applied to the treatment of stuttering by altered feedback and operant procedures. In: Fava, E., editor. *Pathology and therapy of speech disorders*. John Benjamins; Amsterdam: 2002. p. 95–118.
- Howell P. Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary Issues in Communicative Sciences and Disorders*. 2004; 39:122–139.
- Howell P. Signs of developmental stuttering up to age eight and at 12 plus. *Clinical Psychology Review*. 2007; 27:287–306. [PubMed: 17156904]
- Howell, P.; Anderson, A.; Lucero, J. Motor timing and fluency. In: Maasen, B.; van Lieshout, PHHM., editors. *Speech Motor Control: New Developments in Basic and Applied Research*. Oxford University Press; Oxford: in press
- Howell P, Sackin S. Timing interference to speech in altered listening conditions. *Journal of the Acoustical Society of America*. 2002; 111:2842–2852. [PubMed: 12083218]
- Kleinow J, Smith A. Potential interactions among linguistic, autonomic, and motor factors in speech. *Developmental Psychobiology*. 2006; 48:275–287. [PubMed: 16617462]
- Kleinow J, Smith A, Ramig LO. Speech motor stability in IPD: Effects of rate and loudness manipulations. *Journal of Speech, Language, and Hearing Research*. 2001; 44:1041–1051.
- Levelt, W. *Speaking: From intention to articulation*. Bradford Books; Cambridge, MA: 1989.
- Lucero JC, Munhall KG, Gracco VL, Ramsay JO. On the registration of time and the patterning of speech movements. *Journal of Speech, Language, and Hearing Research*. 1997; 40:1111–1117.
- Max L, Caruso AJ. Acoustic measures of temporal intervals across speaking rates: Variability of syllable- and phrase-level relative timing. *Journal of Speech, Language and Hearing Research*. 1997; 40:1097–1110.
- McHenry MA. The effect of pacing strategies on the variability of speech movement sequences in dysarthria. *Journal of Speech, Language, and Hearing Research*. 2003; 46:702–710.
- Packman A, Onslow M, Richard F, van Doorn J. Syllabic stress and variability: A model of stuttering. *Clinical Linguistics and Phonetics*. 1996; 10:235–263.
- Ramsay, JO.; Silverman, BW. *Functional Data Analysis*. Springer-Verlag; New York: 1997.
- Riley, GD. *Stuttering Severity Instrument for Children and Adults*. Third Edition. Pro-Ed; Austin, TX: 1994.
- Smith, A.; Goffman, L. Interaction of motor and language factors in the development of speech production. In: Maasen, B.; Kent, R.; Peters, H.; van Lieshout, P.; Hulstijn, W., editors. *Speech*

- motor control in normal and disordered speech. Oxford University Press; Oxford: 2004. p. 227-252.
- Smith A, Goffman L, Zelaznik HN, Ying G, McGillem C. Spatiotemporal stability and the patterning of speech movement sequences. *Experimental Brain Research*. 1995; 104:493–501.
- Smith, A.; Kelly, E. Stuttering: A dynamic, multifactorial model. In: Curlee, R.; Siegel, G., editors. *Nature and Treatment of Stuttering: New Directions*. second edition. Allyn & Bacon; Needham, MA: 1997.
- Smith A, Kleinow J. Kinematic correlates of speaking rate changes in stuttering and normally fluent adults. *Journal of Speech, Language, and Hearing Research*. 2000; 43:521–536.
- Smith A, Zelaznik H. The development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental Psychobiology*. 2004; 45:22–33. [PubMed: 15229873]
- Stromsta C. A spectrographic study of disfluencies labeled as stuttering by parents. *De Therapia Vocis et Loquellae*. 1965; 1:317–320.
- Walsh B, Smith A. Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research*. 2002; 45:1119–1133.
- Walsh B, Smith A, Weber-Fox C. Short-term plasticity in children's speech motor systems. *Developmental Psychobiology*. 2006; 48:660–674. [PubMed: 17111401]
- Ward D, Arnfield S. Linear and nonlinear analysis of the stability of gestural organization in speech movement sequences. *Journal of Speech, Language, and Hearing Research*. 2001; 44:108–117.
- Wohlert AB, Smith A. Developmental change in variability of lip muscle activity during speech. *Journal of Speech Language and Hearing Research*. 2002; 45:1077–1087.

Figure 1.

A complete set of data for one participant in the experiment showing the steps in computation of STI scores. STI scores for lips (L-STI) in units of mm of displacement are at the top and for envelope (E-STI) in units of sound pressure level (SPL) are at the bottom.

For each signal, the raw scores are top left. The panel to the right represents results after amplitude normalization (L and E). The plots at bottom left for each signal are the results after time normalization. The *sd* across time is presented to the right (overall STI score is indicated in this panel).

Figure 2. Scatter plot of the individual speakers on axes E-STI (abscissa) and L-STI (ordinate). Each point represents a speaker. Speaker-groups can be identified by the symbol (crosses for fluent speakers and diamonds for speakers who stutter).

Figure 3.

Scatter plot of the individual speakers on axes age (abscissa) and L-STI (ordinate). Each point represents a speaker. Speaker-groups can be identified by the symbol (crosses for fluent speakers and diamonds for speakers who stutter).

Figure 4. Scatter plot of the individual speakers on axes age (abscissa) and E-STI (ordinate). Each point represents a speaker. Speaker-groups can be identified by the symbol (crosses for fluent speakers and diamonds for speakers who stutter).

Figure 5.

Mean E-STI (abscissa) and corresponding L-STI (ordinate) for fluent speakers (crosses) and speakers who stutter (diamonds). For each STI signal and for each fluency group, mean STI is given and +/- one standard error (E on abscissa and L on ordinate).

Table 1

Information about individuals in the study. Data about speakers who stutter are given at the top with rows indicating participant number, age, gender, SSI-3 score and its severity designation. Data about control children are given at the bottom with rows indicating participant number, age and gender

Speakers who stutter				
Participant	Age	Gender	SSI-3 Score	Severity
S1	8y 6m	M	31	Severe
S2	10y 0m	M	19	Mild
S3	11y 10m	M	25	Moderate
S4	12y 4m	M	28	Severe
S5	13y 2m	M	22	Moderate
S6	13y 2m	M	16	Mild
S7	14y 2m	M	15	Mild
S8	14y 2m	F	31	Severe
S9	14y 6m	M	10	Very Mild
S10	15y 9m	M	14	Mild
S11	17y 4m	F	25	Moderate
S12	19y 2m	M	18	Moderate
S13	21y 7m	M	9	Very Mild

Controls		
Participant	Age	Gender
C1	8y 0M	F
C2	8y 0m	F
C3	8y 11m	F
C4	11y 5m	M
C5	16y 1m	M
C6	17y 0m	F
C7	17y 6m	F
C8	21y 6m	F
C9	21y 9m	M
C10	23y 3m	M
C11	23y 4m	F
C12	26y 9m	M