JOURNAL ARTICLE

Recording and analyzing kinematic data in children and adults with SOLLAR: Sonographic & Optical Linguo-Labial Articulation Recording system

Aude Noiray[1,2], Jan Ries[1], Mark Tiede[2], Elina Rubertus[1], Catherine Laporte[3] and Lucie Ménard[4,5]

[1] Linguistics department, University of Potsdam, DE

[2] Haskins Laboratories, US

[3] Department of Electrical Engineering, École de technologie supérieure, CA

[4] Département de linguistique, Laboratoire de Phonétique, UQAM, CA

[5] Center for Research on Brain, Language, and Music, CA

Corresponding author: Aude Noiray (anoiray@uni-potsdam.de)

1992

Understanding the development of spoken language in young children has become increasingly important for advancing basic theories of language acquisition and for clinical practice. However, such a goal requires refined measurements of speech *articulation* (e.g., from the tongue), which are difficult to obtain from young children. In recent years though, technological advances have allowed developmental researchers to make significant steps in that direction. For instance, movements of the tongue, an articulator that is essential for spoken language, can now be tracked and recorded in children with ultrasound imaging. This technique has opened novel research avenues in (a)typical language acquisition, enabling researchers to reliably capture what has long remained invisible in the speech of young children. Within this context, we have designed an experimental platform for the recording and the processing of kinematic data: SOLLAR (*Sonographic and Optical Linguo-Labial Articulatory Recording system*). The method has been tailored for children, but it is suitable for adults. In the present article, we introduce the recording environment developed to record over 100 children and 30 adults within SOLLAR. We then describe SOLLAR's data processing framework, providing examples of data visualization and a summary of strengths and limitations.

1. Introduction

In the past decade, empirical research in the developmental domain has benefited from increasingly sophisticated methods for investigating the attention, perception, and recognition abilities of children and infants (e.g., EEG, fNIRs, eye-movement tracking, pupillometry). However, similar methods allowing in-depth examination of the motor mechanisms underpinning spoken language have lagged, due to the invasiveness of the methods necessary to quantitatively measure speech motor activity and/or long pre-recording steps. Collecting kinematic data from children's speech articulators (e.g., from the lips and the tongue) has become increasingly important for fundamental and clinical developmental research because empirical questions remain that cannot be addressed solely via measures of the output of the speech production system (e.g., via formant frequency estimation in vowels) or via the perceptual judgment from expert ears (e.g., phonetic transcriptions). For instance, vowels are often assumed to be mastered by the

age of three (e.g., Kuhl & Meltzoff, 1996), but research has in fact shown that only their production in very simplistic forms is acquired by that age and that some variability persists (e.g., James, Van Doorn, & McLeod, 2001; Lee, Potamianos, & Narayanan, 1999; Ménard, Schwartz, Boë, & Aubin, 2007, Noiray, Cathiard, Abry, & Ménard, 2010).

A fair number of kinematic studies investigating the temporal and spatial organization of children's labial activity (e.g., Noiray, Cathiard, Ménard, & Abry, 2008a; Noiray et al., 2010; Smith & Goffman, 1998; Goffman, Smith, Heisler, & Ho, 2008) and its coordination with the jaw (e.g., Munhall & Jones, 1998; Green, Moore, Higashikawa, & Steeve, 2000; Green, Moore, & Reilly, 2002) has shed light on the maturation of labial control for spoken language fluency. Similar emphasis is needed with respect to the tongue articulator, which, unlike the lips, is invisible yet essential for speaking any language fluently. However, few of the methods employed for tracking adults' lingual activity can be easily optimized for children. For instance, articulography (EMA), which uses small wire-connected sensors glued to the tongue, is commonly used in speech production research to observe lingual activity (Perkell et al., 1992). While the method has been used to describe numerous speech phenomena in adults (see Rebernik, Jacobi, Jonkers, Noiray, Wieling's systematic review, in progress for this collection), it is not well-suited for young children due to necessarily extended preparation times and the inherently invasive nature of the technique. To our knowledge, EMA has hence only been used with school-aged children (e.g., Terband, Maassen, Van Lieshout, & Nijland, 2011). Electropalatography (EPG) was more readily adapted to child speech research (e.g., Gibbon, Hardcastle, & Dent, 1995; Wood, Timmins, Wishart, Hardcastle, Cleland, 2019; Gibbon & Lee, 2017; Gibbon, 1999). Contrary to EMA, EPG estimates places of contact between the tongue and the hard palate (for a comparison between EMA and EPG methods across a single speech dataset produced by adults, see Kochetov, 2020). Because it only requires that an artificial palate be positioned in children's mouths, the method is easy to use a priori. However, EPG is relatively expensive because most systems require a custom-fit palate for each child (and at each visit in the case of longitudinal studies) and even for each adult participant because they do not have uniform palate shapes (e.g., McGarr, Tsunoda, & Harris, 2005).

Since the late 1990's, ultrasound tongue imaging (UTI) has become an increasingly popular technique for studying speech articulation in adults (to only cite a few: Fabre, Hueber, Girin, Alameda-Pineda, & Badin, 2017; Gick, Bird, & Wilson, 2005; Hueber et al., 2010; Kavitskaya, Iskarous, Noiray, & Proctor, 2008; Noiray, Iskarous, & Whalen, 2008b; Wrench & Scobbie, 2011; Stone, 2005; Whalen et al., 2005; Zharkova, 2007). UTI does not track tongue contact (EPG) or tongue flesh points (EMA); instead, an ultrasound probe placed below the speaker's chin enables online tongue surface shape imaging (e.g., using a midsagittal view). Being non-invasive, UTI has gradually been used with typically developing children starting from the age of two years to late puberty (e.g., Barbier et al., 2020; Lenoci & Ricci, 2018; Ménard & Noiray, 2011; Noiray, Ménard, & Iskarous, 2013; Noiray, Abakarova, Rubertus, Krüger, & Tiede, 2018; Rubertus & Noiray, 2020; Noiray, Wieling, Abakarova, Rubertus, & Tiede, 2019a; Rubertus & Noiray, 2018; Song, Demuth, Shattuck-Hufnagel, & Ménard, 2013; Zharkova, Hewlett, & Hardcastle, 2011; Zharkova, Hewlett, & Hardcastle, 2012; Zharkova, 2017) as well as for the description of speech sound disorders (e.g., Bacsfalvi, Bernhardt, & Gick, 2007; Bacsfalvi & Bernhardt, 2011; Bernhardt, Gick, Bacsfalvi, & Adler-Bock, 2005; McAllister Byun, Buchwald, & Mizoguchi, 2016). It has further been implemented as a biofeedback method for the assessment and treatment of speech-related difficulties (e.g., Byun et al., 2014; Cleland, Scobbie, & Wrench, 2015, Cleland, Scobbie, Roxburgh, Heyde, & Wrench, 2017; Cleland, Scobbie, Roxburgh, Heyde, & Wrench, 2019; Preston, Leece, & Maas, 2016; Preston, Leece, & Storto, 2019; Sungden, Lloyd, Lam, & Cleland, 2019). For further information on the topic, we

recommend Sugden, Lloyd, and Cleland's (2019) recent systematic review of clinically oriented ultrasound imaging studies. Last, UTI has been optimized for infant research, e.g., for tracking six- to twelve-month-old infants' communicative tongue movement (Sander, Höhle, & Noiray, 2019), for investigating links between the perception and production of language-specific speech gestures (Bruderer, Danielson, Kandhadai, & Werker, 2015) or for elucidating developmental interactions between speech motor control, lexical, and phonological developments (Noiray et al., 2019b).

While UTI is certainly the most suitable technique for recording kinematic data in children, it also has its drawbacks. First, because it is not designed for speech-related research but borrowed from the medical field, additional devices are often required for recording the acoustic speech signal (e.g., microphone, mixer), keeping the ultrasound probe in a fixed position (as opposed to allowing freehand scanning in the medical field) and storing data (e.g., hard drive, server, computer). Second, before summarizing the ultrasound video data in a way that is amenable to statistical analysis, several time-consuming data processing steps are often needed (e.g., data formatting, tongue contour detection, correction of erroneously generated tongue contours). Importantly for the success of any developmental study, the ultrasound device must be introduced into a child-friendly protocol and, preferably, be operated by experimenters with experience in child research to minimize experimental constraints (e.g., ultrasound gel, sitting still for a long period of time, keeping children focused).

In this context, we have designed a platform dedicated to the recording and processing of child speech called SOLLAR: Sonographic and Optical Linguo-Labial Articulation Recording system. SOLLAR platform uses a spaceship motif to stimulate children's interest in the studies conducted in our laboratory. It allows for the simultaneous recording of the audio speech signal via a microphone, tongue movement via UTI, and lip movement via video recording. The platform has been validated in several studies with children starting from three years of age (Noiray et al., 2018, Noiray et al., 2019a, Noiray et al., 2019b; Rubertus & Noiray, 2018; Rubertus & Noiray, 2020) as well as with adults (Abakarova, Iskarous, & Noiray, 2018). In the remainder of this article, we make suggestions for designing a child-friendly recording environment and describe the data collection protocol developed within the SOLLAR platform (Section 2). We then describe the tongue data processing framework used in our recent studies with German children and adults (Section 3) and provide some examples of tongue data visualization (Section 4). Last, we discuss SOLLAR's strengths and limitations (Section 5).

2. Recordings within SOLLAR

2.1. Creating a child-friendly recording environment

Collecting data sets large enough to anticipate subsequent data exclusion (e.g., due to technical problems, child's inattention, attrition) and enable reliable statistical analyses is especially challenging in all aspects related to child studies (e.g., kinematic, perception, neuroimaging). Researchers must develop creative protocols and/or make substantial efforts when connecting with children to stimulate their attention and interest for the (often monotonous) experimental speech tasks. In the last decade, authors of the present article have conducted various kinematic studies with young children (Ménard & Noiray, 2011; Ménard, Prémont, Trudeau-Fissette, Turgeon, & Tiede, 2020; Noiray, Ménard, Cathiard, Abry, & Savariaux, 2004; Noiray et al., 2010; Noiray et al., 2013; Song, Demuth, & Shattuck-Hufnagel, 2012; Turgeon, Trudeau-Fissette, Fitpatrick, & Meénard, 2017). Leveraging those experiences, we have designed our most recent studies as imaginary interstellar journeys during which child participants pilot a mock spaceship integrated within the SOLLAR platform. The spaceship includes a car seat with seatbelts and measurement tools that resemble those used in airplane cockpits. The small ultrasound

probe is integrated within the control panel of the spaceship. Children are instructed to position their chin on the probe holder so they can take off and undertake the planned interstellar journey. With this approach, children understand that the probe is a crucial component of the spaceship like a gas pedal in cars and are willing to stay still in order to complete the space journey.

To stimulate children's attention, our storyline merges aspects of gaming and storytelling. During the imaginary interstellar journeys, children travel to six planets, complete a series of missions to enable traveling to the next planet (i.e., the speech-related production tasks) and take pictures of the newly encountered alien friends. With this storyline, we aimed to 1) provide the children with a visual timeline indicating their progress in the task, 2) create an impression of movement and hence compensate for the need to sit relatively still in the lab for half an hour. While this scenario would be unrealistic from an adult perspective, it worked very well for over 100 children recorded in our lab (Noiray et al., 2018; Noiray et al., 2019a, Rubertus & Noiray, 2018; Rubertus & Noiray, 2020). Upon sitting in the spaceship, children choose an avatar from a set of small puppets. The puppet is then placed in a miniature spaceship, taped on a sidewall on the planet Earth, the starting point of their journey. Six other planets are shown on the wall to probe the six randomized lists of stimuli planned in one of our studies. Before leaving Earth and upon returning to Earth after reaching the last alien planet, children drink water with a straw while positioned on the probe to acquire images of their palate. Hence, our storyline, like most children's stories, includes beginning and end points (the Earth), characters (their avatar, aliens to be met on each planet), a set of actions (missions to be completed between each planet, i.e., in our case repeating or reading lists of words), and regular rewards (stickers of alien pictures to be taped in a customized booklet). The booklet helps children remain focused and motivated in completing the speech-related tasks. We adapt the pictures and booklet to the children's ages. We noticed that when they reach the first year of primary school, many children want to be treated more like adults and become offended if they perceive the pictures as childish. For additional motivation, children are promised and awarded a stamped certificate and space-themed present (e.g., a space-themed jigsaw puzzle) if they completed all missions (i.e., the study). Last, in consideration of children's limited attention spans, we make sure that recordings do not exceed 40 minutes including introduction of the study, set-up, and breaks.

2.2. Role of experimenters in child recordings

While creating a child-friendly environment may facilitate the collection of quantitative kinematic data from young children, experimenters' patience and engagement are the best motivations for children. To create a friendly connection between experimenters and child participants, we use various strategies.

In our studies, responsibilities have been dispatched between two experimenters: a participant relations experimenter (PE) and a desk experimenter (DE). Upon arrival at the lab, the PE describes the study to the adult participant or, in the case of a child participant, to both parents and the child, collects written consent, and helps participants or their parents fill in various questionnaires. The PE is the only experimenter to interact with the participant during the experimental phase; that is, she/he is in charge of the familiarization period, testing preparation, and the speech production tasks. The desk experimenter (DE) instead operates all devices, which are hidden behind screens to avoid distracting participants during testing. She/he also controls for the quality of the data collected (e.g., participant position, video, and ultrasound image quality).

Before the data collection starts, the PE engages with the child familiarizing him/her with the ultrasound device in a playful way, explaining the space mission's goals with excitement,

asking what they know about planets and how they feel about the space adventure. Because the child needs to wear goggles for subsequent pixels to mm conversion (see Section 3.4) and apply blue markers on their face to correct for possible head movement during post-processing (see Section 3.2), the PE may also wear goggles and applies markers on his/her own face to connect with the child and create an empathetic atmosphere. During the testing period, she/he regularly encourages the child in completing the missions and monitors their comfort. Pauses are made after the completion of each interplanetary flight (i.e., production of a predetermined list of words). During those breaks, the PE talks about the aliens with the child and gives her/him positive feedback.

2.3. Equipment used for SOLLAR's recording platform

SOLLAR is a multimodal recording platform that supports concurrent recordings of speech audio using a directional microphone (Sennheiser), tongue movement using a portable ultrasound imaging device (Sonosite Edge, 48Hz), labial-shape variation, and head motion using a video camera (Sony, 60fps). All devices are integrated into the SOLLAR spaceship motif. The microphone is attached to the spaceship control panel. The small ultrasound probe is integrated within a custom-made probe holder positioned below the participant's chin to image the tongue surface contour on the midsagittal plane. The probe holder restrains movement of the ultrasound probe to vertical translation only, to track jaw movement (**Figure 1**). It is mounted in an adjustable custom-made pedestal that is fully integrated to the spaceship. The length of the pedestal can be adjusted manually depending on space constraints and experimental requirements. The pedestal is positioned on an adjustable electrical table to allow larger variation in height.

During the recording, children are comfortably seated in an armchair suitable for children. It includes seatbelts and it is tilted slightly upwards in the front for the legs to remain stable. Participants are instructed to remain still and look at a bright star positioned above the camera in front of them while keeping their chin on the ultrasound probe. The PE stands behind the star to keep constant eye contact with the child or adult participant and operates the presentation of the speech stimuli using a laptop computer. Adult participants instead sit on a larger armchair.

Simultaneous views of the front and profile of the participant's face can be obtained via a mirror positioned at a 45° angle, reflecting the participant's profile into the video camera's field of view. Alternatively, we have also used two separate webcams positioned in the
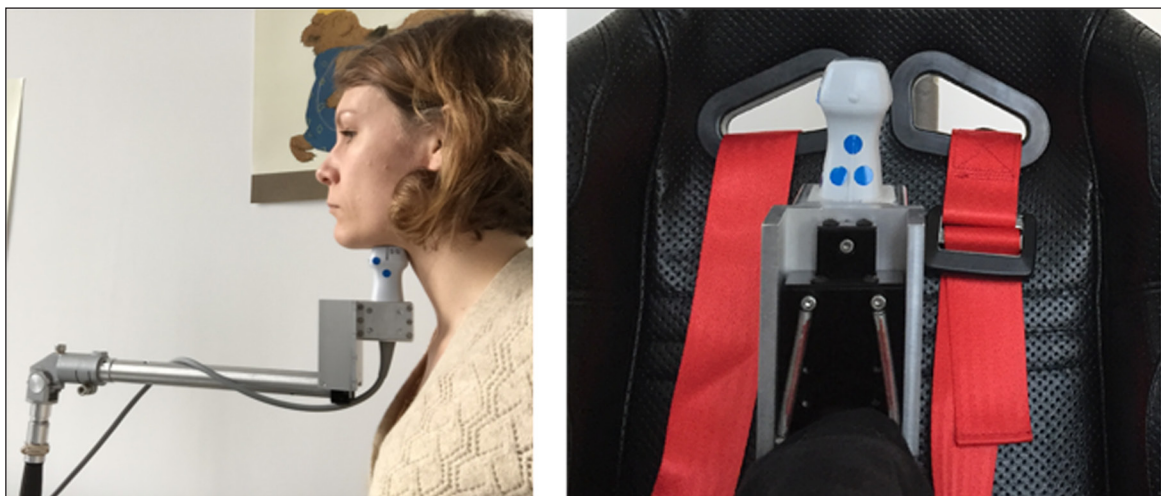

Figure 1: Left: Profile view of the ultrasound probe, probe holder, and pedestal when not integrated into the spaceshift's control panel, right: front view of the probe, probe holder, and car seat.

front and at the side of the participant to get simultaneous face and profile views without the mirror (see **Figure 2**). Video is digitized using an AverMedia GameBroadcaster HD video capture card, which combines the ultrasound video stream with the audio signal from the microphone into a single video recording using the Open Broadcaster Software Studio (OBS, http://obsproject.com). The camcorder video is captured by a Blackmagic Design Intensity Shuttle video interface and also recorded using OBS. In the dual-webcam set-up, the two video streams are combined in a split-view image, and we use the audio stream from the frontal camera for synchronization of the two video streams. The UTI and video recording streams are synchronized offline after the recording, by maximizing the cross-correlation of their respective audio signals. For this, we use MATLAB's cross-correlation function to compute the time lag between the streams (e.g., in adults: Abakarova et al., 2018; Noiray, Cathiard, Ménard, & Abry, 2011; Noiray, Iskarous, & Whalen, 2014; in children: Noiray, et al., 2010; Noiray, et al., 2018; Noiray, et al., 2019a; Rubertus & Noiray, 2018). See Section 3 for a full description of the process.

2.4. *Strategies for probe and head stabilization*

Because UTI is adversely affected by movement of the head or probe away from the optimal midsagittal view, several strategies have been developed to minimize and correct for head movement. The choice of strategy depends on the target population, type of data to be collected, tolerance to invasiveness, and cost effectiveness. Current strategies include a headgear that maintains the probe in a relatively fixed position (e.g., Zharkova et al., 2011), head mounted probe stands (e.g., Barbier et al., 2020), and motion tracking approaches that relate the position of the head and probe for post-recording correction (e.g., HOCUS; Whalen et al., 2005, used with children in Ménard et al., 2020). In general, increased constraints on head motion lead to more constrained and thus less natural speech, since the head and jaw cannot move freely with respect to the ultrasound probe; however, the resulting tongue positions can be analyzed without extensive post-processing. Conversely, unconstrained approaches require reconciliation of the time-varying spatial relationship between the head and probe for analysis, and excessive translation or torsion of the head with respect to the probe may make the recording unusable. For children, constraining head motion is invasive and potentially intimidating, so a frequently used approach is to hold the probe in place by hand instead (e.g., Ménard & Noiray, 2011; Zharkova, Gibbon, & Hardcastle, 2015). However, if the participant moves a lot, this may result in substantial probe movement or inconsistent contact between the probe and the
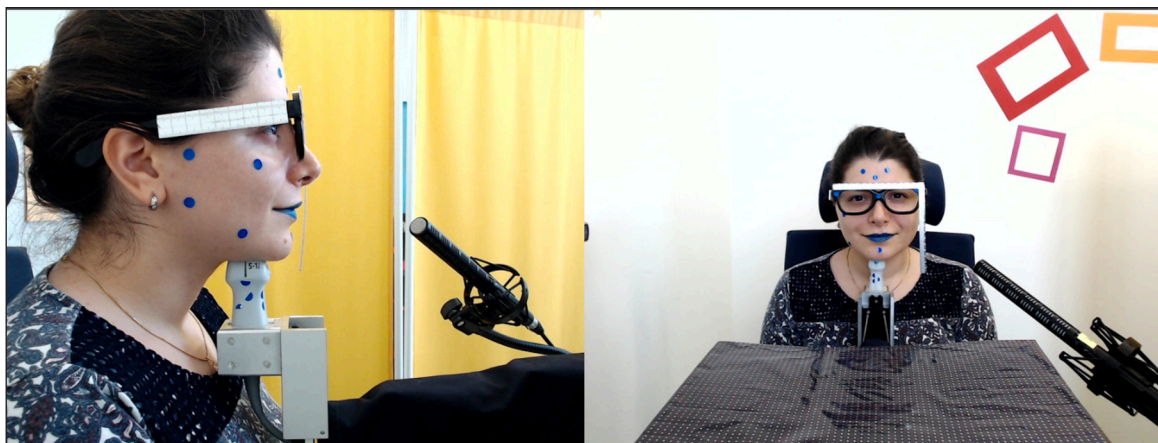

Figure 2: Profile (left) and frontal (right) views on the face of an adult speaker during a recording with SOLLAR.

chin, which in turn may greatly affect the quality of the ultrasound images collected. The results are also uncalibrated with respect to palatal hard structure.

With SOLLAR, we have developed an approach that avoids intimidating head restraint and minimizes deleterious movement, while accommodating the vertical jaw displacements associated with normal speech (**Figure 2**). The customized probe holder allows movement only along the vertical axis. To support our video-based head tracking, we apply a series of small adhesive blue markers to the participant's face, about 5 mm in diameter (see **Figure 2**). Scaling (mm/pixels) is provided by a spectacle frame with marked rulers attached to its front and sides. Blue markers are also attached to the front and side of the ultrasound probe to track its position relative to the head. While there is some amount of flexibility in the tracking procedure, the general arrangement includes:

- Three markers on participant's forehead: one marker centered slightly above the eyebrows and two more markers set above and to the left and right of the first marker;
- Four markers on the right side of participant's face: one on the zygomatic bone underneath the eye, one on the temple close to the ear, one close to the angle of the mandible, and one on the mandible bone close to the mouth opening;
- One marker on the chin;
- Three markers each on the front and side of the ultrasound probe, arranged in a triangular shape.

This results in four sets of markers—i.e., head and ultrasound probe both in frontal and profile views—that are subsequently used to match positions across recorded stimuli blocks and track motion frame-by-frame. The triangular configuration of each set of markers allows measurements of displacement along the x- and y-axes as well as some rotations. With the simultaneously recorded frontal and profile views we are able to estimate head movements corresponding to neck flexion in both left-right and dorsal-ventral directions, with the former appearing mainly as an artefact and the latter occurring during natural speech. Left-right head rotation is not considered, but participants are instructed to face forward during the experiment. Marker tracking and motion correction in SollarSuite is a multi-step process and is described in greater detail in Section 3.3.

3. Description of SollarSuite

3.1. General description

Once all of the different types of raw data are recorded, they need to be synchronized and processed to correct for head movement, extract tongue surface contours, and so on. These steps can be time-consuming. To address this, the SollarSuite package of data processing and analysis tools was developed in MATLAB (see **Figure 3** for a flowchart showing its main components). SollarSync synchronizes the different raw data streams by cross-correlating the audio streams, thus creating a common timecode and building a frame-by-frame table incorporating keyframes identified in the acoustic labeling. Tongue contours are traced in SollarContours and stored in the integrated data structure. In a second step, head motion tracking is performed using the video data. For this, a reference frame is defined for each participant and the configuration of blue markers on the head and probe in both the frontal and profile views are taken as tracking templates. Each recorded experimental block is matched to this reference frame; within-block movement is estimated by a frame-by-frame point tracking algorithm (**Figure 5**, Section 3.4).

With this two-step tracking procedure of across-block matching and within-block point tracking, a combined transformation matrix can be computed for each frame, representing
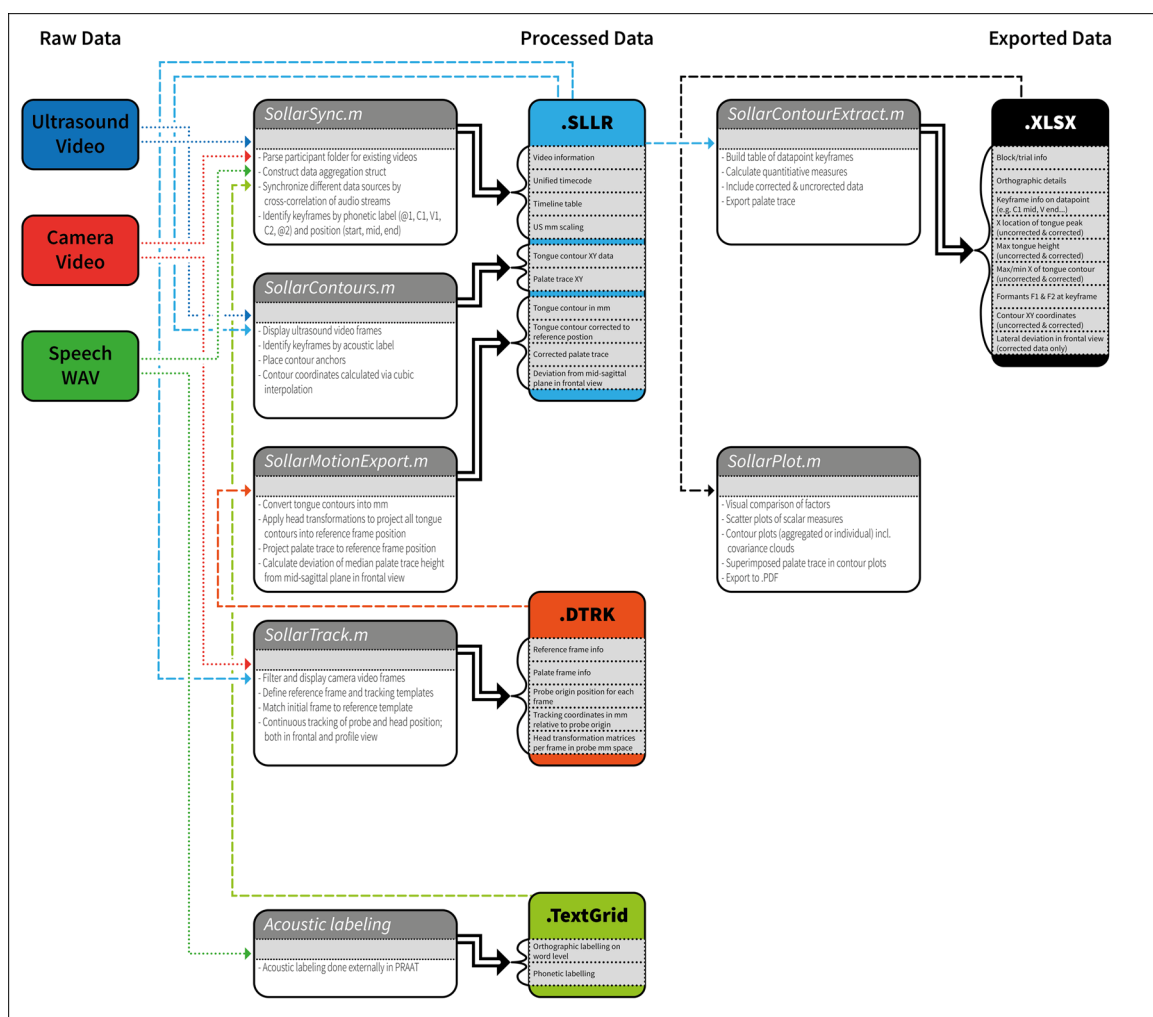
Figure 3: Flowchart highlighting the main components of SollarSuite and how they integrate the different data streams.

the rigid transformation necessary to correct the difference in head position relative to the ultrasound probe's point of origin to the spatial configuration of the reference frame. For the profile view, this transformation can be applied to the ultrasound contour trace, thus a) correcting for variation introduced by head movement (see Section 3.3) and b) aligning each tongue contour to the hard palate trace recorded separately in the swallow recordings (see Section 3). In the frontal view, the lateral displacement of the head along the probe surface can be quantified and a threshold for discarding single trials can be applied. All information is integrated into the common data structure and can be inspected using SollarPlot or extracted using SollarContourExtract.

3.2. Data processing (SollarSynch)

As a first step in pre-processing the raw data, the SollarSync.m tool is used to synchronize the different data sources and create a data structure that is used to pass data between the different components of SollarSuite. As a prerequisite, SollarSync expects raw data to be placed in one folder per subject, with subfolders US, WAV, PRAAT, and CAM containing the raw data files of the different data sources. Recordings from these data sources are matched by filename, i.e., a recording of one block or session is represented by an identically named file in each folder. Not all data sources are mandatory and SollarSync provides fallback options for missing data:

- US data is the mandatory core data for SollarSuite and available ultrasound video files in this folder are taken as the basis to look for other data sources. Cannot be missing;
- WAV contains high-quality voice recordings of the participants. These files are usually the basis for any acoustic labelling done in PRAAT (Boersma & Weenink, 2016) and the synchronized timeline is created from them. If no corresponding wave files are found, SollarSync extracts the audio stream from the US videos as a fallback;
- PRAAT TextGrid files are optional. If found, all available tiers are imported to the data structure and are available for keyframe selection in further processing;
- CAM contains video files from an external camera. In the SOLLAR setup, these recordings combine frontal and profile views of the participant's head with tracking markers applied to the face to allow for head motion tracking. If no such video data is recorded, a version of SollarSuite that does not attempt motion tracking is available separately.

SollarSync will run on such a structured data folder without further user interaction and with status information displayed in MATLAB's Command Window (release R2019a). A matrix of all data sources found will be shown first and the synchronization process will proceed through this list with the following steps:

- US and CAM video files are analyzed for duration, picture size, and frame rate;
- The audio streams from WAV, US, and CAM sources are cross-correlated to estimate the lag between each source. The estimated lag reflects the fact that different streams are recorded with slightly different starting times. The voice audio recording is taken as the reference stream for synchronization and lag values for other sources are computed relative to the audio recording starting point. These lag values are then added to each frame's individual timestamp, resulting in a timecode table for each data source where a certain time point in the audio recording can be associated with the corresponding frame in each video stream;
- Lastly, available PRAAT TextGrid files are parsed for SollarSync to import all available intervals or point tiers into its data structure. SollarSync also calculates points of interest for all interval tiers, namely beginning and end points, midpoint, as well as a point three-quarters through the interval. Those are added as point tiers, with four points per interval with added suffixes _000, _050, _075, and _100.

3.3. Tongue surface contour detection (SollarContours)

Analysis and synchronization results are stored in a structured data array and saved as an .sllr file within the participant folder. For tongue contour tracing, SollarSuite builds upon GetContours, (https://github.com/mktiede/GetContours), a Matlab-based program for fitting discretized tongue surface contours to ultrasound imaging data (Tiede & Whalen, 2015). The program supports image preprocessing, sequence playback, and frame selection. Click-and-drag positioning of reference points control a cubic spline fit to the currently displayed image frame, which can then be refined using an integrated active contour model ('snake').

SollarContours.m contains an extended GUI that ties in with the SollarSuite by making use of the .sllr data structure (**Figure 4**).

When a recording is loaded, SollarContours offers any tier information found in the data structure as a source for defining keyframes. By default, it identifies any labelled frame as a keyframe, but frames can be manually selected and deselected. A timeline view at the bottom of the GUI window indicates keyframes and whether a tongue contour trace
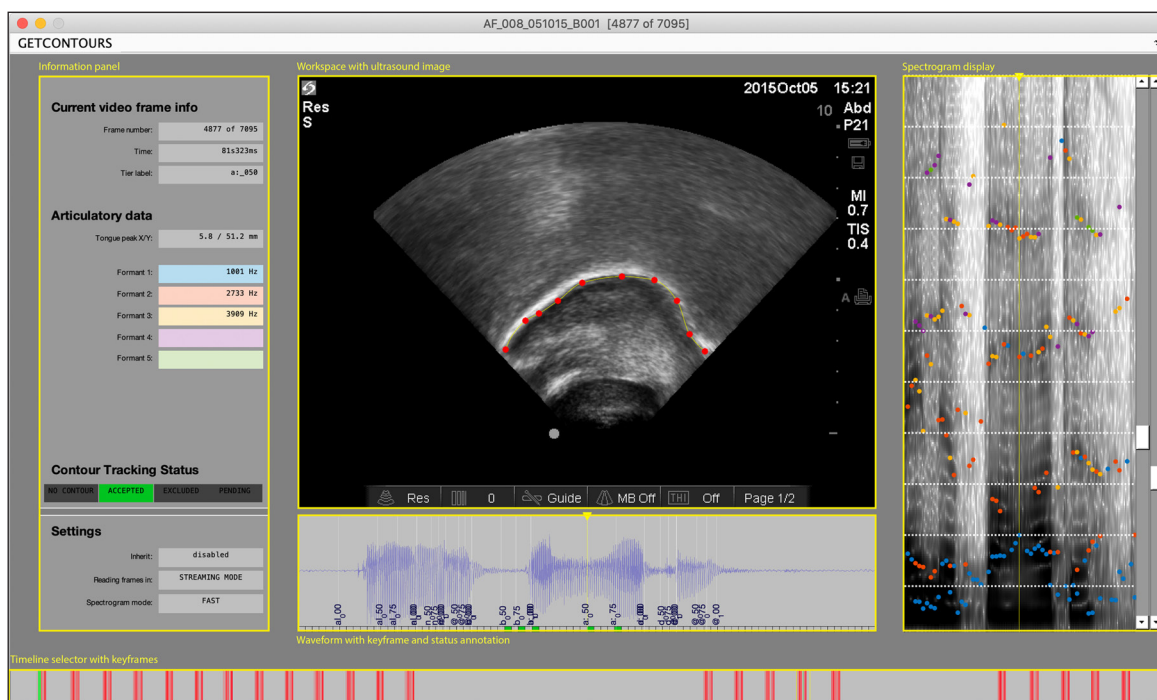
Figure 4: Display of SollarContours.

is found or not. This timeline view also serves as a way to quickly navigate the data. The currently selected video frame is displayed in the central workspace and framed by graphs and panels displaying additional information: The info panel on the left presents the current frame number and time, any labels attached to this frame, as well as formant and tongue contour parameters; a spectrogram view on the right visualizes the acoustic signal in the currently selected time segment, and a waveform of the current segment is presented below, including keyframe labels and status indicators.

The central workspace provides the functionality known from GetContours. In GetContours, tongue tracing is performed by placing anchors on the ultrasound image and a continuous contour is interpolated between the anchor points. Click-and-drag positioning of reference points control a cubic spline fit to the currently displayed image frame, which can then be refined using an integrated active contour model ('snake'). The majority of GetContours' features are preserved in SollarContours, e.g., redistributing anchors, inheriting anchors between frames, image filtering, as well as the navigational tools and associated keyboard commands. They are accessible in SollarContours' menu bar.

In addition to manual contour tracing, SollarContours allows the import of tongue contours generated with SLURP (Laporte & Ménard, 2018). SLURP is a publicly available MATLAB-based software tool for automatically tracing tongue contours in ultrasound video data. Given a small number of anchor points manually positioned on any single frame of the video, SLURP uses a particle filtering method to robustly track an active contour (Li, Kambhamettu, & Stone, 2005) across the video in a compact space parameterized by contour location, length, and a small number of shape characteristics. Automatic SLURP tracking is performed outside SollarSuite and the resulting .mat file can be selected for import from a menu item. Any contour data found in this file is transformed into the SollarContours format and merged into the data structure, including the energy map as a measure of contour quality. In the process of importing SLURP data, the contour coordinates are transformed into an anchor-based representation. Specifically, a stepwise approximation of the tongue contour is calculated with an incrementally increasing

number of anchor points until the sum of absolute differences between the interpolated and original tongue contour falls below a threshold. This way, the contour trace is imported into the SollarContours workspace in a sparse and conveniently editable format, and integrates seamlessly with the known functionality of GetContours.

For additional data quality control, SollarContours introduces a status variable for each frame, flagging it as accepted, excluded, or pending. Freshly imported tongue contour data is by default labelled as pending, i.e., awaiting manual confirmation. Keyframe navigation and keyboard shortcuts are implemented in SollarContours to reduce the time required for this process.

3.4. Head and probe movement correction (SollarTrack)

SollarSuite includes SollarTrack.m, a GUI-based tool to administer and monitor motion tracking of head and probe (**Figure 5**). It integrates with the other components of SollarSuite through use of the previously discussed data structure contained in a participant's .sllr file. It also relies on the same fixed folder structure and initialization with SollarSync.m has to be performed first. Tracking results are stored separately in a TRACKING subfolder and as a .dtrk file. Head and probe motion tracking can be performed almost independently from contour tracing, except that the ultrasound frame containing the hard palate trace needs to be specified and should thus be identified with SollarContours beforehand.

Markers tracking begins by finding and setting the reference frame, which will be the baseline position for each video recording made during the participant's testing session. This is of special importance to the frontal head position, as the position in the reference frame is taken as the zero point when it comes to head displacement or left-right neck flexion in relation to the ultrasound probe. Consequently, it is important to select a frame that depicts the participant in a relaxed posture, with the head upright and just above the ultrasound probe. The participant should also not be articulating, show no strong facial expressions so as to not affect the relative positions of the blue markers (e.g., frowning), and all blue markers should be clearly and entirely visible.

Tracking templates are defined for the reference frame by sequentially selecting blue markers for each of the four tracking sets: frontal head view, profile head view, frontal
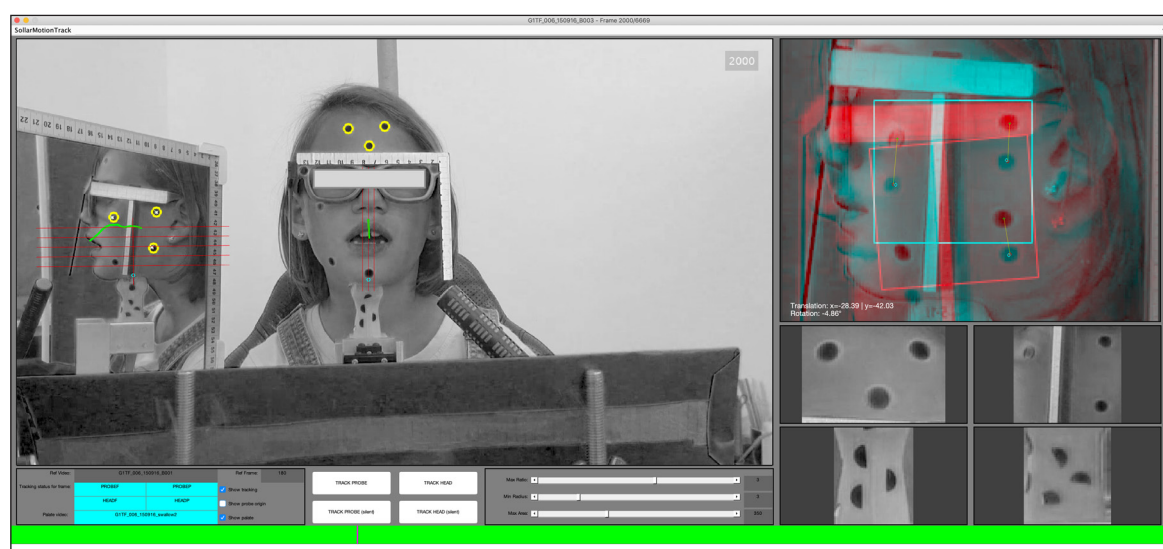

Figure 5: Screenshot of the SollarTrack GUI, displaying a frame of the camera video with tracking markers (yellow circles) and palate trace (green lines) superimposed. Smaller panels on the right-hand side show template matching information (top) and the four reference templates.

probe view, profile probe view. SollarTrack puts no constraints on the number or layout of markers for each template, but a triangular configuration has proven robust for computing the spatial translations. When choosing a marker template, a good visibility of all chosen markers throughout the whole recording should be considered. SollarTrack automatically identifies blue markers by isolating pixels that fall within the blue color range and by applying inclusion criteria, such as minimum and maximum size and roundness, to the candidate regions in the resulting binary image. In the selection process, SollarTrack takes the centroid coordinates of a selected marker to relieve the user of pixel-perfect precision and to ensure more accurate matching across video recordings in the following step.

For the probe templates, two additional specifications are necessary: First, SollarTrack will ask the user to draw a Probe Orientation Line starting from the probe origin and extending downwards to capture the angle of the probe within the camera image. Second, a 5 cm segment has to be selected on the scales attached to the goggles. This measure is used to compute the conversion factor from pixels to mm. For continuous tracking in a recorded video, each of the four tracking templates in the starting frame of the video are matched to their locations in the reference frame. The rigid transformation between the two frames is estimated[1] and stored within the tracking data structure and its corresponding .dtrk file. Frame-by-frame tracking is then performed on probe templates in a two-step process of tracking marker positions first, then computing the rigid transformation for each frame. Working from the binary image, SollarTrack takes all pixels that fall within the vicinity of the selected blue markers. These pixel coordinates are fed into a MATLAB PointTracker object and SollarTrack proceeds stepwise through the frames. Single pixels with invalid tracking results are disregarded in all further frames, which is compensated for by the large number of initial pixels but could be a limitation when tracking long recordings. After point tracking is complete, the rigid transformation is estimated between one frame and the next and progressively combined into a transformation matrix that reflects rigid motion between each frame and the reference frame.

Consecutive rigid transformations are then applied to compute the coordinates of the probe origin for each frame, both for frontal and profile views. Including probe orientation and pixel-to-mm-ratio information from the reference frame, mm-based coordinates with respect to the probe origin are calculated for frontal and profile head templates. This means that head position can now be referred to in the same coordinate system as a mm-corrected tongue contour trace. Subsequent frame-by-frame tracking of both head templates follows the same two-pass process as probe template tracking, but additionally includes estimation of the rigid transformations in the mm-based coordinate system.

This mm-based tracking information represents changes in the spatial relationship between the rigid structure of the participants' heads to the probe origin. It is consequently used to align tongue contour data in a common space, including the hard palate trace. It improves data quality by removing the variability introduced by head motion, either by correcting for changes in the head's position over the ultrasound probe as calculated from the profile view or by removing single trial data, when frontal view head motion data suggest a substantial deviation from a midsagittal ultrasound view of the tongue.

---

[1] For geometric estimation only translation and rotation are being considered, as SollarTrack is performing under the assumption that the motion of the blue markers in their respective frontal/profile view is reasonably well represented by a flat, rigid structure moving on a two-dimensional plane. Movements that violate this assumption, such as head rotation or large changes in distance to the camera, should only occur when the participant exhibits uninstructed behavior and would be excluded as artifacts.

3.5. Data exploration and export

Lastly, SollarSuite offers tools for data inspection, visualization, and export. These tools help identify potentially remaining artifacts for manual exclusion and provide a convenient way to visualize and explore the data before exporting it for statistical analysis. While some of these features can be used independently, we present them here through the SollarPlot.m data visualization tool as it incorporates all of these features.

SollarPlot reads and aggregates data from multiple subjects, structures them by defining different data types, and gives the user the option to produce either scatter plots of extracted scalar data (e.g., the x-position of the tongue apex) or whole tongue contour plots. In each case, different conditions and factors can be applied to separate the data within a plot (as separate lines) or into individual plots. In its current state, the extraction routines are tailored towards extracting a specific set of data points relating to the kind of keyframe labelling we use in our studies and would have to be adapted for use in other studies. The different kind of data types applied in SollarPlot are:

- datapoint: this signifies that a column in the tabular data indicates different points of interest within one trial, e.g., vowel midpoint
- data: labels a column as containing scalar, numeric data available as a source for the scatter plot functionality. Examples are points of minimum or maximum height of the tongue and also formants
- contour: labels the content of a column as contour data available for source selection in contour plots
- factor: columns labelled this way are being offered for selection when separating the data into individual plots

An additional functionality within SollarPlot allows for the creation of new factor variables, in which the values of existing variables can be mapped onto new values. For example, if Block exists as a factor variable, this could be coded again into the first and second half of the experiment by mapping the block numbers onto values 1 and 2 within a newly created dummy variable.

When using the scatter plot functionality, a data variable is selected as the source of the data to be plotted as well as a datapoint variable for the selection of points of interest. All unique values found in the datapoint variable can be selected for the x- and y-axis independently. The resulting scatter plot will show how for each trial a tongue contour parameter relates between points of interest. **Figure 6** illustrates this with the tongue body position on the front-back dimension as a function of the highest point on the tongue body as data source. The datapoint variable here is the temporally segmented phonetic labelling tier from which two time points of interest are selected: the consonant midpoint (C1_050) on the y-axis and vowel midpoint (V_050) on the x-axis. Separated by color are the three consonants /b, d, g/ in various vocalic contexts. Each point in this scatter plot therefore shows how the front-back value of the tongue body relates between consonant and vowel midpoints across all consonant-vowels trials.

This visualization can be further restricted by selecting factor variables to either plot data with different factor values as separate colors within one plot and/or to plot them in separate axes. The result will be displayed in a scatter plot including a regression line, with the r^2 value indicated in the plot's legend. The details of the selection are also printed to MATLAB's Command Window as well as a plain-text log file in the current working directory, along with the parameters of the linear regression such as n, degrees of freedom, r^2, and slope.
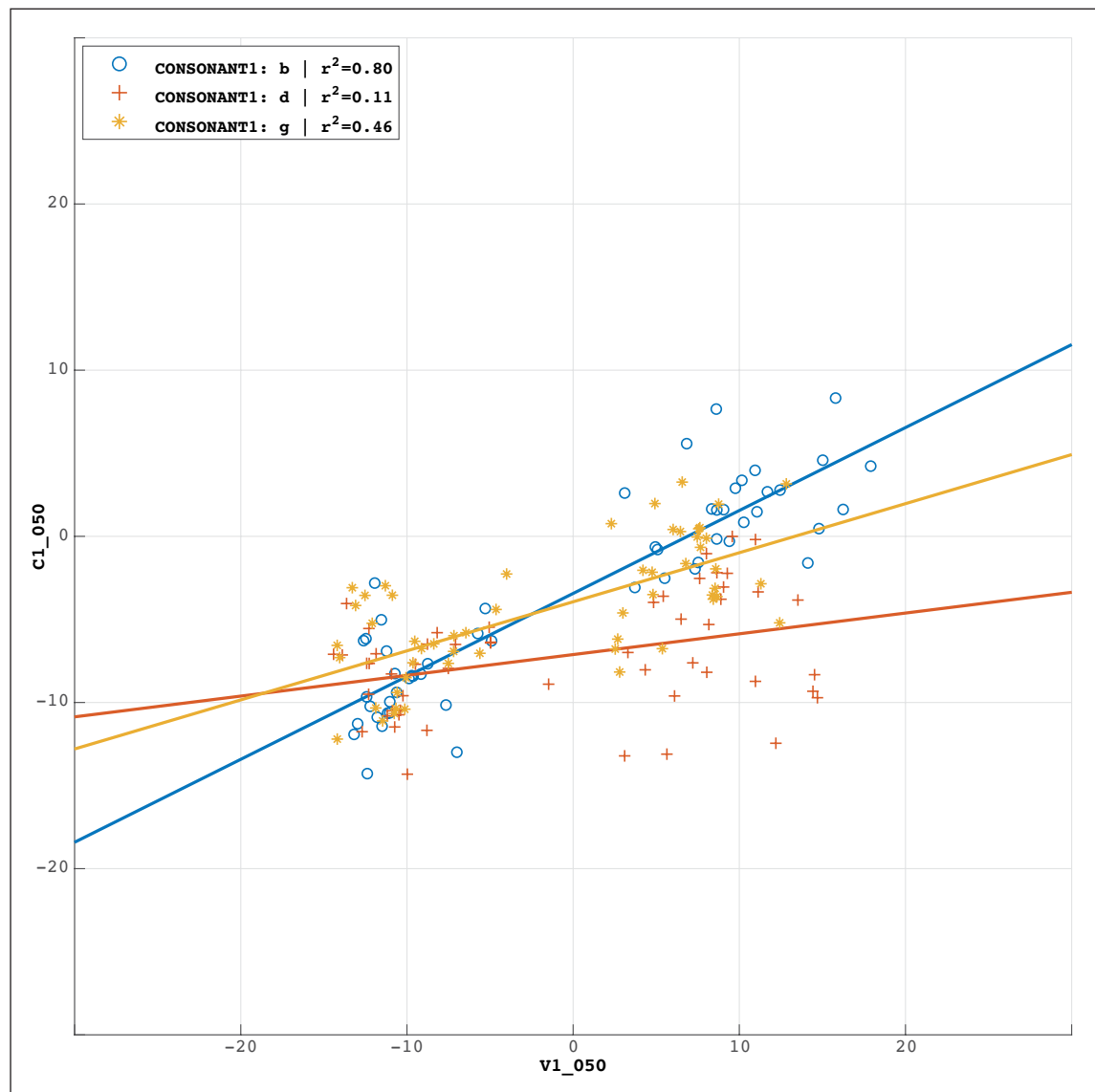
Figure 6: Scatter plot illustrating the front-back tongue body value position for the highest point on the tongue body at the temporal midpoint of a vowel (x-axis) and previous consonant (y-axis) in CV sequences produced repeatedly by an adult speaker.

Similarly, for the contour plot functionality, a contour variable is selected as the source and again a datapoint variable for the selection of points of interest. From the values in the latter, one point of interest is selected and an averaged tongue contour plot is created, that again can be broken down further by separating according to factor variable values. For the resulting display, range or covariance clouds can be revealed in the plots, allowing visualization of the variability of tongue contours that enter the averaging. To identify outliers, the individual contours can also be added to the plot and reveal block and trial information when selected by the user.

Data aggregation and export is performed upon loading data into SollarPlot and files for further analysis are stored automatically during that process. They include tabular data per subject in .xlsx Excel spreadsheet format within each included participant's folder. In this spreadsheet, each row corresponds to a point of interest within one trial with data and factors as columns. Additionally, a MATLAB .mat file is stored in the current working directory which includes the same data as the .xlsx files but for all participants that have been selected for aggregation within SollarPlot. This file is also used for data storage

within SollarPlot, which means that it is updated as dummy variables are created within SollarPlot and can be loaded again when re-opening SollarPlot at a later time. In addition to that, an 'Export Plots' button in SollarPlot's GUI is available for both kinds of plots. It creates a clean rendition of the current plot in a separate figure window, prompts the user for a filename, and exports this plot to PNG and EPS formats.

4. Examples of data visualization

In the following, we showcase applications of SollarSuite to demonstrate how it can help pre-process ultrasound data and improve the quality of tongue contours, especially when dealing with a recording situation wherein less control is exercised over the participants (e.g., with children). The studies mentioned below have been approved by the ethical Committee of the University of Potsdam and conform to the Declaration of Helsinki.

The first example (**Figure 7**) provides an illustration of six averaged midsagittal tongue contours of an adult participant created with SollarPlot. These tongue contours were obtained subsequent to the recording of pseudo-words elicited for a study investigating coarticulatory effects from vowels onto various preceding consonants. This example includes all elicitations of one participant of the velar stop /g/ averaged separately for
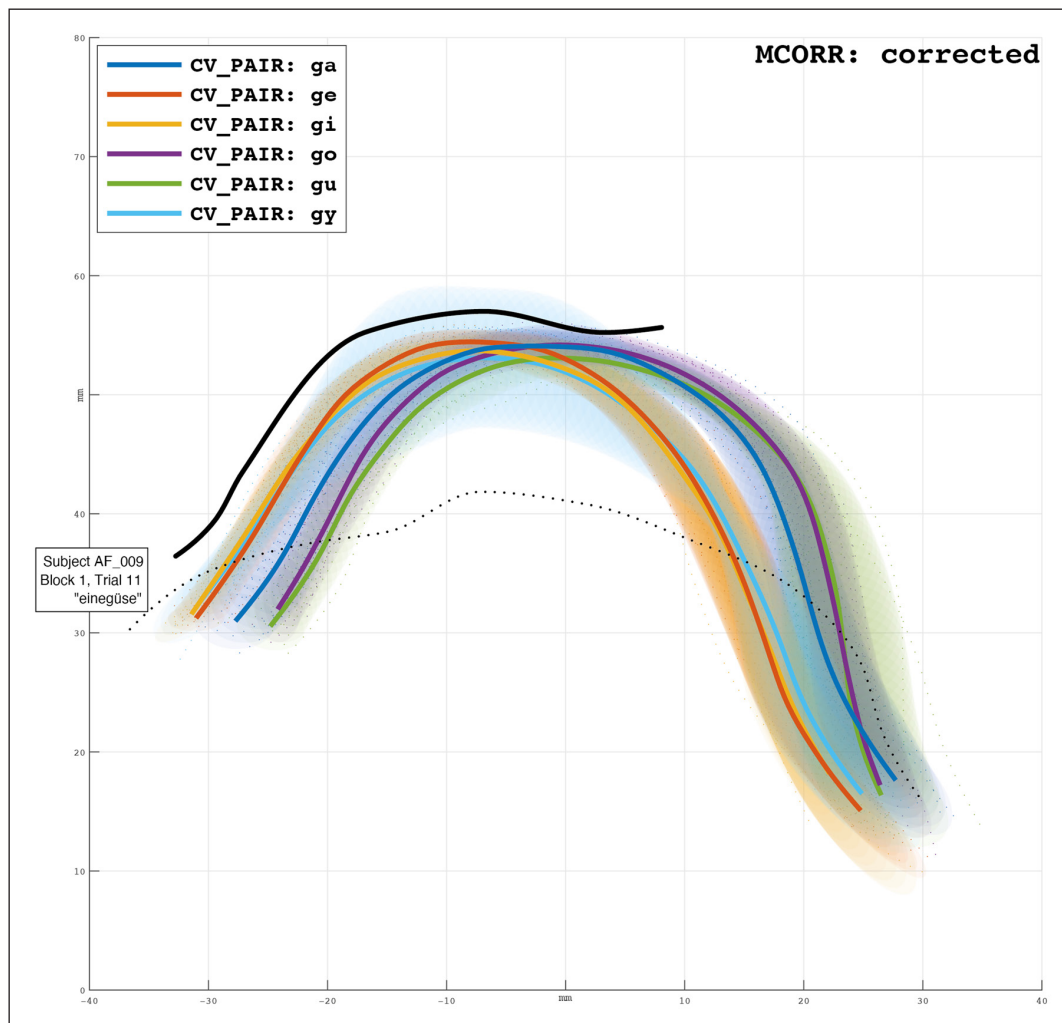

Figure 7: Averaged midsagittal tongue contours of an adult speaker created with SollarPlot. Left side: anterior part of the tongue; right side: back of the tongue. Each colored tongue contour represents the temporal midpoint of /g/ in CV syllables with various vocalic contexts. The black line illustrates an estimate of the hard palate structure.

six different following vowels (/i/, /a/, /u/, /y/, /e/, /o/). The tongue contours are selected at the temporal midpoint of the acoustically-defined velar stop that includes the consonant closure and burst. These elicitations were recorded over several blocks and into separate video files. In addition to that, a series of water bolus images were recorded in another video to obtain a trace of the hard palate, which is represented by the thick black line in **Figure 7**. The resulting ultrasound and camera recordings were synchronized using SollarSync, tongue contours were manually traced for keyframes using SollarContours, and head position was continuously tracked using SollarTrack. The head position information was then used to align all tongue traces in the same coordinate system, resulting in the combined plot of averaged tongue shapes in relation to the hard palate.

Variability within the averaged tongue contours is visualized in **Figure 7** as covariance clouds. In this example, we can see a large variability for the elicitations of /g/ in /gy/ sequences (in light blue). To identify the source of this variability, we used SollarPlot to show all individual tongue contours in this plot. We were then able to identify the outlier, which is shown in **Figure 7** by a dotted black line along with an information box with trial specifics. This points us to the specific block and trial number, which we can now manually inspect to identify the source of this digression. In this case, a review of the camera video revealed that a labelling mistake misidentified the uttered pseudo-word as /gyzə/ while it was actually /zygə/.

As outlined above, we designed the SOLLAR setup specifically with child participants in mind. While our adult participants typically report no problems following our request to keep a straight, forward-facing posture over the course of a recording session, a lot more movement is expected with young children. Hence, in the child cohorts, the application of SollarTrack is especially important—not only to align data from separate blocks with the hard palate trace, but also to account for differences in head position with respect to the probe. We apply a correction to account for head movement when possible, and use exclusion criteria in cases where a reliable recording of the midsagittal view of the tongue cannot be guaranteed.

Figure 8 provides an illustration of this application in contour plots created with SollarPlot with the data collected from a seven-year-old participant. The plots depict averaged tongue contours of the stops /b/, /d/, /g/ during the temporal midpoint of the acoustically defined domain of the consonant and the fricative /z/ in CV syllables, with the left panel containing uncorrected midsagittal contours and the right panel containing the same data after motion correction was applied and rejected trials excluded. The covariance cloud visualization indicates a reduced variability within the same consonant. The effect of correction can be seen distinctly when looking at the tip of the corrected tongue contours for the alveolar consonants in this set, /d/ and /z/, where the tongue position is quite narrowly prescribed during articulation. However, even after correction, a substantial amount of variability can still be exhibited in the tongue blade and dorsum for those two consonants, as well as for /b/ and /g/, for which a larger degree of coarticulation may take place (Noiray et al., 2019a). This indicates that applying SollarTrack's motion correction specifically removes variability introduced by head movement, while natural variability in tongue motion during speech production is preserved.

5. Summary of strengths and limitations

In the past years, the SOLLAR platform has allowed us to collect kinematic data to investigate coarticulatory mechanisms in over 100 children from three to nine years of age (e.g., Noiray et al., 2019a; Noiray et al., 2018; Rubertus & Noiray, 2018; Rubertus & Noiray, 2020) as well as over 30 adults (Abakarova et al., 2018), and further examine aloud
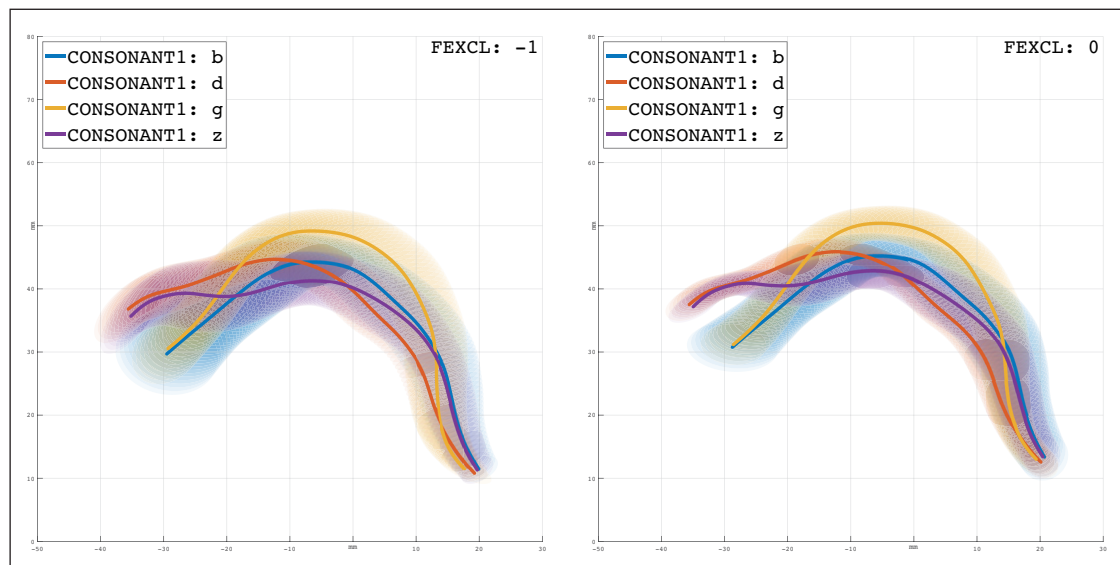
Figure 8: Midsagittal tongue contours at the midpoints of four consonants for a seven-year-old child. Left: highly variable contours prior to motion correction. Right: reduced variability subsequent to corrective transformations applied and trials excluded, for which the head is displaced more than 5 mm laterally above the probe.

reading fluency in 30 children in primary school (Popescu & Noiray, 2020). In addition to tracking the tongue, SOLLAR is designed for future integration of a labial shape tracking system inspired from previous research conducted at the GIPSA Lab (e.g., with adults: Lallouache, 1991; Noiray et al., 2011; Ménard, Leclerc, & Tiede, 2014; Sodoyer, Rivet, Girin, Savariaux, Schwartz, & Jutten, 2009; with children Noiray et al., 2010). During the production tasks, participants' lips can be painted in blue as this color maximizes contrast with the skin. In post-processing the video data, the blue shapes corresponding to the lips can be tracked for measurement of lip aperture, interlabial area, and upper lip protrusion. While this feature could easily be integrated in SollarSuite, it has not been our focus so far.

In the future, SOLLAR can potentially be used in clinical practice, e.g., for the description and diagnostic of speech-related disorders (e.g., speech sound disorder: Cleland et al., 2015; stuttering: Lenoci & Ricci, 2018). However, in its current state, the SOLLAR platform requires some space and uses several pieces of equipment in addition to the ultrasound device which may only be available in laboratories, not in speech and language therapy offices. In such conditions, one may want to use a more compact set-up (e.g., Cleland, Wrench, Lloyd, & Sugden, 2018).

To consider space limitations, **Table 1** summarizes the main strengths, limitations, and perspectives for improvement for each component included in SOLLAR.

6. Conclusion

SOLLAR has been designed to respond to the growing need among developmental psycholinguists and phoneticians to collect kinematic data in young children and the concurrent lack of suitable methods. While SOLLAR does not solve all experimental challenges, it has been designed as a child-friendly environment that can fairly easily be implemented to record kinematic data in young children. In future studies, it may be combined with other behavioral methods (e.g., eye-movement tracking, EEG) to develop more integrated empirical approaches to language acquisition, in which concurrent examinations of speech motor and cognitive abilities are possible (e.g., perception, attention) as well as their on-line interactions.

Table 1: Summary table of SOLLAR's strengths, limitations, and perspectives for improvements.

	Strengths	Limitations & known Problems	Perspectives for improvement
Recording platform	– Child-friendly, maintains children's interest and motivation to complete the task	– two experimenters needed to conduct the study (one operating all devices, the other to monitor children) – Multiple devices needed in addition to the ultrasound device	– Using research-oriented ultrasound devices that includes synchronized audio signal recording, substantial storage of high-quality ultrasound video and potentially includes a video camera
Video Camera setup	– Inexpensive: two webcams and blue stickers	– USB cameras and recording software not built for accurate synchronization; most likely only correct to within one or two frames – Trade-off needs to be found between video file size, image quality, and frame-exact image retrieval when applying video codec settings – Requires dedicated video recording machine with sufficient power	– Replacing individually placed markers with larger tracking marker could help calibrate for mm distance, replacing the obtrusive spectacles – Separately recording the two webcams and using audio stream synchronization could improve temporal accuracy, but complicates lab setup
SollarSync	– Data synchronization by cross-correlating audio is very reliable – Shared .sllr data structure removes the need for dealing with a multitude of files	– Video handling in MATLAB depends on the capabilities of the computer and operating system and can produce slightly different results from one platform to the next when building a shared timecode – Storing data in a single structure makes data slightly less accessible	– Greater flexibility and adjustability should be a main direction for future updates – Speed improvements and batch processing capabilities could be added
Sollar-Contours	– Expanded GUI makes main functionality of GetContours more accessible to researchers less familiar with MATLAB – Import of SLURP data can speed up tongue contour tracing by relying on automatically generated data – Additional navigational functionality in GUI improves its use as a tool for data inspection	– Tongue detection (or manual correction) is time consuming – Forked off of a previous version of GetContours, i.e., updates to GetContours have to be manually ported to SollarContours – Displays a fair amount of currently unused information, e.g, spectrogram with formants – Some assumptions about our specific setup are hard-coded	– GUI should be reworked to focus on most-used elements – Greater flexibility for differently sourced ultrasound data

(Contd.)

	Strengths	Limitations & known Problems	Perspectives for improvement
SollarTrack	– Good flexibility regarding configuration of tracking templates – Combined approach of reference matching between recordings and frame-by-frame tracking within recording proved a reliable method	– Chance of increasingly unreliable point tracking when videos are long – Performs time consuming pre-import of video frames (to then greatly speed up tracking and motion calculation) – Requires relatively large amounts of memory – Relies on MATLAB's parallel computing toolbox, which might not be available to all users – Handling of problems tracking should provide better options for user manual intervention	– Feature to exclude partial segments of a video, where tracking is interrupted, in testing stage – Development of tracking algorithms to inform experimental setup and emphasize importance of strict testing protocol

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Abakarova, D., Iskarous, K., & Noiray, A. (2018). Quantifying lingual coarticulation in German using mutual information: An ultrasound study. *The Journal of the Acoustical Society of America, 144*(2), 897–907. DOI: https://doi.org/10.1121/1.5047669

Bacsfalvi, P., & Bernhardt, B. M. (2011). Long-term outcomes of speech therapy for seven adolescents with visual feedback technologies: Ultrasound and electropalatography. *Clinical Linguistics & Phonetics, 25*(11–12), 1034–1043. DOI: https://doi.org/10.3109/02699206.2011.618236

Bacsfalvi, P., Bernhardt, B. M., & Gick, B. (2007). Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *Advances in Speech Language Pathology, 9*(1), 36–45. DOI: https://doi.org/10.1080/14417040601101037

Barbier, G., Perrier, P., Payan, Y., Tiede, M. K., Gerber, S., Perkell, J. S., & Ménard, L. (2020). What anticipatory coarticulation in children tells us about speech motor control maturity. *Plos one, 15*(4), e0231484. DOI: https://doi.org/10.1371/journal.pone.0231484

Bernhardt, B., Gick, B., Bacsfalvi, P., & Adler-Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics, 19*(6–7), 605–617. DOI: https://doi.org/10.1080/02699200500114028

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (Version 6.0.20) Available from http://www.praat.org/

Bruderer, A. G., Danielson, D. K., Kandhadai, P., & Werker, J. F. (2015). Sensorimotor influences on speech perception in infancy. *Proceedings of the National Academy of Sciences, 112*(44), 13531–13536. DOI: https://doi.org/10.1073/pnas.1508631112

Byun, T. M., Hitchcock, E. R., & Swartz, M. T. (2014). Retroflex versus bunched in treatment for rhotic misarticulation: Evidence from ultrasound biofeedback intervention. *Journal of Speech, Language, and Hearing Research, 57*(6), 2116–2130. DOI: https://doi.org/10.1044/2014_JSLHR-S-14-0034

Cleland, J., Scobbie, J., Roxburgh, Z., Heyde, C., & Wrench, A. A. (2017). Ultraphonix: Using ultrasound visual biofeedback to teach children with special speech sound disorders new articulations. In *7th International Conference on Speech Motor Control*.

Cleland, J., Scobbie, J. M., Roxburgh, Z., Heyde, C., & Wrench, A. (2019). Enabling new articulatory gestures in children with persistent speech sound disorders using ultrasound visual biofeedback. *Journal of Speech, Language, and Hearing Research, 62*(2), 229–246. DOI: https://doi.org/10.1044/2018_JSLHR-S-17-0360

Cleland, J., Scobbie, J. M., & Wrench, A. A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical linguistics & phonetics, 29*(8–10), 575–597. DOI: https://doi.org/10.3109/02699206.2015.1016188

Cleland, J., Wrench, A., Lloyd, S., & Sugden, E. (2018). ULTRAX2020: Ultrasound Technology for Optimising the Treatment of Speech Disorders: Clinicians' Resource Manual. Glasgow: University of Strathclyde. DOI: https://doi.org/10.15129/63372

Fabre, D., Hueber, T., Girin, L., Alameda-Pineda, X., & Badin, P. (2017). Automatic animation of an articulatory tongue model from ultrasound images of the vocal tract. *Speech Communication, 93,* 63–75. DOI: https://doi.org/10.1016/j.specom.2017.08.002

Gibbon, F. E., Hardcastle, W., & Dent, H. (1995). A study of obstruent sounds in school-age children with speech disorders using electropalatography. *International Journal of Language and Communication Disorders, 30*(2), 213–225. DOI: https://doi.org/10.3109/13682829509082532

Gibbon, F. E. (1999). Undifferentiated lingual gestures in children with articulation/phonological disorders. *J Speech Lang Hear Res. Apr; 42*(2), 382–97. DOI: https://doi.org/10.1044/jslhr.4202.382. PMID: 10229454.

Gibbon, F. E., & Lee, A. (2017). Electropalatographic (EPG) evidence of covert contrasts in disordered speech. *Clinical linguistics & Phonetics, 31*(1), 4–20. DOI: https://doi.org/10.1080/02699206.2016.1174739

Gick, B., Bird, S., & Wilson, I. (2005). Techniques for field application of lingual ultrasound imaging. *Clinical Linguistics & Phonetics, 19*(6–7), 503–514. DOI: https://doi.org/10.1080/02699200500113590

Goffman, L., Smith, A., Heisler, L., & Ho, M. (2008). The breadth of coarticulatory units in children and adults. *Journal of Speech, Language, and Hearing Research, 51,* 1424–1437. DOI: https://doi.org/10.1044/1092-4388(2008/07-0020)

Green, J. R., Moore, C. A., Higashikawa, M., & Steeve, R. W. (2000). The physiologic development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research, 43*(1), 239–255. DOI: https://doi.org/10.1044/jslhr.4301.239

Green, J. R., Moore, C. A., & Reilly, K. J. (2002). The sequential development of jaw and lip control for speech. *Journal of Speech, Language, and Hearing Research, 45,* 66–79. DOI: https://doi.org/10.1044/1092-4388(2002/005)

Hueber, T., Benaroya, E. L., Chollet, G., Denby, B., Dreyfus, G., & Stone, M. (2010). Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips. *Speech Communication, 52*(4), 288–300. DOI: https://doi.org/10.1016/j.specom.2009.11.004

James, D., Van Doorn, J., & McLeod, S. (2001). Vowel production in mono-, di-and poly-syllabi words in children 3;0 to 7;11 years. In *Speech Pathology Australia National Conference: Evidence and Innovation* (pp. 127–136). Speech Pathology Australia. DOI: https://doi.org/10.3109/14417040109003718

Kavitskaya D., Iskarous K., Noiray, A., & Proctor, M. (2008). Trills and palatalization: Consequences for sound change. *Proceedings of the 17th Meeting of Formal Approaches to Slavic Linguistics*, New Haven, May 9–11, pp. 97–110.

Kochetov, A. (2020). Research methods in articulatory phonetics I: Introduction and studying oral gestures. *Language and Linguistics Compass, 14*(4), 1–1. DOI: https://doi.org/10.1111/lnc3.12368

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America, 100*(4), 2425–2438. DOI: https://doi.org/10.1121/1.417951

Lallouache, M. T. (1991). *Un poste "visage-parole" couleur: Acquisition et traitement automatique des contours des lèvres.* Doctoral dissertation, Grenoble INPG.

Laporte, C., & Ménard, L. (2018). Multi-hypothesis tracking of the tongue surface in ultrasound video recordings of normal and impaired speech. *Medical image analysis, 44,* 98–114. DOI: https://doi.org/10.1016/j.media.2017.12.003

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. DOI: https://doi.org/10.1121/1.426686

Lenoci, G., & Ricci, I. (2018). An ultrasound investigation of the speech motor skills of stuttering Italian children. *Clinical linguistics & phonetics*, *32*(12), 1126–1144. DOI: https://doi.org/10.1080/02699206.2018.1510983

Li, M., Kambhamettu, C., & Stone, M. (2005). Tongue motion averaging from contour sequences. *Clinical linguistics & phonetics*, *19*(6–7), 515–528. DOI: https://doi.org/10.1080/02699200500113863

McAllister Byun, T., Buchwald, A., & Mizoguchi, A. (2016). Covert contrast in velar fronting: An acoustic and ultrasound study. *Clinical linguistics & phonetics*, *30*(3–5), 249–276. DOI: https://doi.org/10.3109/02699206.2015.1056884

McGarr, N. S., Tsunoda, K., & Harris, K. S. (2005). Palatography: A comparison between custom-made and "flexible" artificial palates for speech production measures. *Journal of the Acoustical Society of America*, 86(S1): 1989. DOI: https://doi.org/10.1121/1.2027316

Ménard, L., Leclerc, A., & Tiede, M. (2014). Articulatory and acoustic correlates of contrastive focus in congenitally blind adults and sighted adults. *Journal of Speech, Language, and Hearing Research*, *57*(3), 793–804. DOI: https://doi.org/10.1044/2014_JSLHR-S-12-0395

Ménard, L., & Noiray, A. (2011). The development of lingual gestures in speech: Experimental approach to language development. *Faits de langues*, *37*, 189–202. DOI: https://doi.org/10.1163/19589514-037-01-900000011

Ménard, L., Prémont, A., Trudeau-Fissette, P., Turgeon, C., & Tiede, M. (2020). Probing the development of phonemic goals in French through prosodic focus. *Journal of Speech, Language, and Hearing Research*.

Ménard, L., Schwartz, J. L., Boë, L. J., & Aubin, J. (2007). Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics, 35*(1), 1–19. DOI: https://doi.org/10.1016/j.wocn.2006.01.003

Munhall, K. G., & Jones, J. A. (1998). Articulatory evidence for syllabic structure. *Behavioral and Brain Sciences*, *21*(4), 524–525. DOI: https://doi.org/10.1017/S0140525X98391268

Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. *Journal of Speech, Language, and Hearing Research, 61*(6), 1355–1368. DOI: https://doi.org/10.1044/2018_JSLHR-S-17-0148

Noiray, A., Cathiard, M. A., Abry, C., & Ménard, L. (2010). Lip rounding anticipatory control: Cross linguistically lawful and ontogenetically attuned. *Speech motor control: New developments in basic and applied research*, 153–171. DOI: https://doi.org/10.1093/acprof:oso/9780199235797.003.0009

Noiray, A., Cathiard, M. A., Ménard, L., & Abry, C. (2008a). Emergence of a vocalic gesture control: Attunement of the anticipatory rounding temporal pattern in French children. In S. Kern, F. Gayraud & E. Marsico (Eds.), *Emergence of language Abilities* (pp. 100–116). Cambridge Scholars Publishing.

Noiray, A., Cathiard, M. A., Ménard, L., & Abry, C. (2011). Test of the movement expansion model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *The Journal of the Acoustical Society of America, 129*(1), 340–349. DOI: https://doi.org/10.1121/1.3518452

Noiray, A., Iskarous, K., & Whalen, D. H. (2008b). Tongue-jaw synergy in vowel height production: Evidence from American English. In R. Sock, S. Fuchs & Y. Laprie, (Eds.), *Proceedings of 8th International Speech Production Seminar, Strasbourg*. Strasbourg, France, pp. 81–84.

Noiray, A., Iskarous, K., & Whalen, D. H. (2014). Variability in English vowels is comparable in articulation and acoustics. *Laboratory phonology*, *5*(2), 271–288. DOI: https://doi.org/10.1515/lp-2014-0010

Noiray, A., Ménard, L., Cathiard, M. A., Abry, C., & Savariaux, C. (2004). The development of anticipatory labial coarticulation in French: A pionering study. In *Eighth International Conference on Spoken Language Processing.*

Noiray, A., Ménard, L., & Iskarous, K. (2013). The development of motor synergies in children: Ultrasound and acoustic measurements. *The Journal of the Acoustical Society of America*, *133*(1), 444–452. DOI: https://doi.org/10.1121/1.4763983

Noiray, A., Popescu, A., Killmer, H., Rubertus, E., Krüger, S., & Hintermeier, L. (2019b). Spoken language development and the challenge of skill integration. *Frontiers in Psychology*, *10*, 2777. DOI: https://doi.org/10.3389/fpsyg.2019.02777

Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., & Tiede, M. (2019a). Back from the future: Nonlinear anticipation in adults' and children's speech. *Journal of Speech, Language, and Hearing Research*, *62*(8S), 3033–3054. DOI: https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0208

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., & Jackson, M. T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *The Journal of the Acoustical Society of America*, *92*(6), 3078–3096. DOI: https://doi.org/10.1121/1.404204

Popescu, A., & Noiray, A. (2020). Coarticulatory organization in beginner readers: a multifactorial interaction approach. *Proceedings of the International Seminar on Speech Production, Providence*, RI (switched to an online conference).

Preston, J. L., Leece, M. C., & Maas, E. (2016). Intensive treatment with ultrasound visual feedback for speech sound errors in childhood apraxia. *Frontiers in Human Neuroscience*, *10*, 440. DOI: https://doi.org/10.1044/2018_LSHSS-18-0081

Preston, J. L., Leece, M. C., & Storto, J. (2019). Tutorial: Speech motor chaining treatment for school-age children with speech sound disorders. *Language, speech, and hearing services in schools*, *50*(3), 343–355.

Rebernik, T., Jacobi, J., Jonkers, R, Noiray, A., & Wieling, A. (in progress for this collection). Reviewing 30 years of using electromagnetic articulography: Some suggestions for improved experimental approaches. *Laboratory Phonology*. DOI: https://doi.org/10.3389/fnhum.2016.00440

Rubertus, E., & Noiray, A. (2018). On the development of gestural organization: A cross-sectional study of vowel-to-vowel anticipatory coarticulation. *PloS one, 13*(9), e0203562. DOI: https://doi.org/10.5334/labphon.228

Rubertus, E., & Noiray, A. (2020). Vocalic activation width decreases across childhood: Evidence from carryover coarticulation. *Laboratory Phonology*, *11*(1), 7.

Rubertus, E., Popescu, A., & Noiray, A. (2020). Development of coarticulation: Comparing modalites in beginning readers. *Proceeding International Speech Production Seminar*, Providence. DOI: https://doi.org/10.1371/journal.pone.0203562

Sander, J., Höhle, B., & Noiray, A. (2019). From the eye to the mouth: Does a developmental shift in attention co-emerge with the emergence of babbling? *Conference Phonetics and Phonology in Europe.*

Smith, A., & Goffman, L. (1998). Stability and patterning of speech movement sequences in children and adults. *Journal of Speech, Language, and Hearing Research*, *41*(1), 18–30. DOI: https://doi.org/10.1044/jslhr.4101.18

Sodoyer, D., Rivet, B., Girin, L., Savariaux, C., Schwartz, J. L., & Jutten, C. (2009). A study of lip movements during spontaneous dialog and its application to voice activity detection. *The Journal of the Acoustical Society of America*, *125*(2), 1184–1196. DOI: https://doi.org/10.1121/1.3050257

Song, J. Y., Demuth, K., & Shattuck-Hufnagel, S. (2012). The development of acoustic cues to coda contrasts in young children learning American English. *The Journal of the Acoustical Society of America*, *131*(4), 3036–3050. DOI: https://doi.org/10.1121/1.3687467

Song, J. Y., Demuth, K., Shattuck-Hufnagel, S., & Ménard, L. (2013). The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound. *Journal of Phonetics*, *41*(3–4), 281–295. DOI: https://doi.org/10.1016/j.wocn.2013.03.004

Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical linguistics & phonetics*, *19*(6–7), 455–501. DOI: https://doi.org/10.1080/02699200500113558

Sugden E., Lloyd, S., Lam, J., & Cleland, J. (2019). Systematic review of ultrasound visual biofeedback in intervention for speech sound disorders. *International Journal of Language and Communication Disorders*, *54*, 05–728. DOI: https://doi.org/10.1111/1460-6984.12478

Terband, H., Maassen, B., Van Lieshout, P. H. H. M., & Nijland, L. (2011). Stability and composition of functional synergies for speech movements in children with developmental speech disorders. *Journal of Communication Disorders*, *44*(1), 59–74. DOI: https://doi.org/10.1016/j.jcomdis.2010.07.003

Tiede, M., & Whalen, D. H. (2015). GetContours: An interactive tongue surface extraction tool. *Proceedings of Ultrafest VII.*

Turgeon, C., Trudeau-Fissette, P., Fitpatrick, E., & Meénard, L. (2017). Vowel intelligibility in children with cochlear implants: An acoustic and articulatory study. *International Journal of Pediatric Otorhinolaryngology*, *101*, 87–96. DOI: https://doi.org/10.1016/j.ijporl.2017.07.022

Whalen, D. H., Iskarous, K., Tiede, M. K., Ostry, D. J., Lehnert-LeHouillier, H., Vatikiotis-Bateson, E., & Hailey, D. S. (2005). The Haskins optically corrected ultrasound system (HOCUS). *Journal of Speech, Language, and Hearing Research*. DOI: https://doi.org/10.1044/1092-4388(2005/037)

Wood, S. E., Timmins, C., Wishart, J., Hardcastle, W. J., & Cleland, J. (2019). Use of electropalatography in the treatment of speech disorders in children with Down syndrome: A randomized controlled trial. *Int J Lang Commun Disord. Mar; 54*(2): 234–248. DOI: https://doi.org/10.1111/1460-6984.12407. Epub 2018 Jul 24. PMID: 30039902.

Wrench, A. A., & Scobbie, J. M. (2011). Very high frame rate ultrasound tongue imaging. In *Proceedings of the 9th International Seminar on Speech Production (ISSP).*

Zharkova, N., Gibbon, F. E., & Hardcastle, W. J. (2015). Quantifying lingual coarticulation using ultrasound imaging data collected with and without head stabilisation. *Clinical linguistics & phonetics*, *29*(4), 249–265. DOI: https://doi.org/10.3109/02699206.2015.1007528

Zharkova, N. (2007). Quantification of coarticulatory effects in several Scottish English phonemes using ultrasound. *QMU Speech Science Research Centre Working Papers, WP-13.*

Zharkova, N., Gibbon, F. E., & Lee, A. (2017). Using ultrasound tongue imaging to identify covert contrasts in children's speech. *Clinical linguistics & phonetics*, *31*(1), 21–34. DOI: https://doi.org/10.1080/02699206.2016.1180713

Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2011). Coarticulation as an indicator of speech motor control development in children: An ultrasound study. *Motor Control*, *15*(1), 118–140. DOI: https://doi.org/10.1123/mcj.15.1.118

Zharkova, N., Hewlett, N., & Hardcastle, W. J. (2012). An ultrasound study of lingual coarticulation in/s V/syllables produced by adults and typically developing children. *Journal of the International Phonetic Association*, *42*(2), 193–208. DOI: https://doi.org/10.1017/S0025100312000060