The Dynamics of Vowel Hypo- and Hyperarticulation in Swedish Infant-Directed Speech to 12-Month-Olds

Ellen Marklund* and Lisa Gustavsson

Phonetics Laboratory, Stockholm Babylab, Department of Linguistics, Stockholm University, Stockholm, Sweden

Vowel hypo- and hyperarticulation (VHH) was investigated in Swedish infant-directed speech (IDS) to Swedish 12-month-olds using a measure that normalizes across speakers and vowels: the vhh-index. The vhh-index gives the degree of VHH for each individual vowel token, which allows for analysis of the dynamics of VHH within a conversation. Using both the vhh-index and traditional measures of VHH, the degree of VHH was compared between Swedish IDS and ADS. The vowel space area was larger in IDS than in ADS, and the average vhh-index as well as the modal value was higher in IDS than in ADS. Further, the proportion of vowel tokens that were highly hyperarticulated (vhh-index > 75th percentile) were fewer in ADS than in IDS. Vowels in Swedish IDS to 12-month-olds are thus concluded to be hyperarticulated compared to vowels in Swedish ADS, both in terms of degree and frequency. Findings are in line with previous reports on VHH in Swedish IDS as well as on VHH in IDS to infants around 12 months in other languages. The study considers the importance of robust formant estimation, highlights the need for replication of studies on VHH in IDS on previously studied languages and ages, and discusses the benefits of the vhh-index. Those benefits include that it normalizes across speakers and vowels, can be used for dynamic measures within speech samples, and permits analyses on token-level.

Keywords: vowel hypo- and hyperarticulation, VHH, infant-directed speech, IDS, Swedish, vhh-index, hyperarticulation, hypoarticulation

BACKGROUND

Hypo- and Hyperarticulation in Speech

Speech production is highly variable. Depending on a multitude of factors, for example speech rate (Adams et al., 1993) and segmental context (Stevens and House, 1963), the precise articulation of a particular speech sound varies considerably, even within a single speaker. According to the H&H theory (Lindblom, 1990), articulation can be placed on a continuum between clear or exaggerated articulation (hyperarticulation) and more relaxed articulation with more reductions (hypoarticulation). The cause of this variability is the speaker's adaptiveness, where he or she balances the listener's need for a clear signal with preservation of energy in articulatory motor activity. For example, when signal transmission is less than ideal, such as in noisy environments, the speaker tends to hyperarticulate, whereas under optimal transmission conditions the speaker defaults to hypoarticulate because it requires less energy (Lindblom, 1983). The speaker's
adaptiveness to the listener’s need for a clear signal extends to internal factors as well; if the speaker has reason to assume that the listener knows what he or she will say next, he or she tends to hypoarticulate, whereas more unexpected content is more likely to be hyperarticulated (Munson and Solomon, 2004). This in-the-moment adaptation on part of the speaker can result in articulation varying on the hyper-hypo continuum not only between different conversations taking place in different conditions, but also within a single conversation and even a single utterance.

In addition to speaker adaptation, both segmental context (Stevens and House, 1963) and linguistic stress (van Bergem, 1993) contribute to variability in speech sound articulation. For example, provided the rate of articulation is stable and relatively rapid, coarticulation will result in a more hypoarticulated back vowel in a /sVs/ context than in a /pVp/ context, since the tongue may not reach target articulation in the first case due to the articulatory demands of the dental fricative context, whereas in the latter case the tongue has more time to reach target articulation as it can be positioned in preparation for the vowel already during the first bilabial, whose articulation is not dependent on tongue position. In general, slower speech tend to be more hyperarticulated than faster speech, since the tongue has more time to reach target articulation in the former case than in the latter (Moon and Lindblom, 1994). In the same vein, stressed words and syllables typically have longer duration than unstressed words and syllables (Fant and Kruckenberg, 1994), and are, as a result, also typically more hyperarticulated (Lindblom, 1963).

Vowel Hypo- and Hyperarticulation in Infant-Directed Speech

Infant-directed speech (IDS) is a speech style that people adopt when talking to young children (for a review, see Soderstrom, 2007). It is characterized both by linguistic simplifications (e.g., Elmlinger et al., 2019), high degree of positive vocal affect (Kitamura and Burnham, 2003), as well as prosodic modifications such as high and variable fundamental frequency (f0) (e.g., Fernald et al., 1989) and longer intrapersonal parental pauses (e.g., Marklund et al., 2015), compared to adult-directed speech (ADS). In addition, the degree to which vowels (e.g., Kuhl et al., 1997; Benders, 2013), consonants (e.g., Sundberg and Lacerda, 1999; Benders et al., 2019) and lexical tones (e.g., Kitamura et al., 2001; Liu et al., 2007) are hypo- and/or hyperarticulated has been shown to differ between IDS and ADS.

Vowel hypo- and hyperarticulation (VHH) in IDS has been investigated in a multitude of studies, focusing on different infant ages, different languages, and different vowels in different contexts, as well as using different measures of VHH. Yet no clear picture emerges; in many studies, IDS vowels have been found to be more hyperarticulated than those in ADS (e.g., Kuhl et al., 1997; Kalashnikova and Burnham, 2018; Kalashnikova et al., 2018), but in others they have instead been found to be more hypoarticulated (e.g., Englund and Behne, 2006; Benders, 2013; Xu Rattanasone et al., 2013). The following review focuses on studies having reported VHH in terms of vowel space area, as this is the most commonly used measure. This measure entails calculating the area in acoustic space (typically F1-F2 space) between the mean formant values of a given set of vowels (typically the point vowels /i/, /a/, and /u/). Findings from studies using other measures are summarized at the end of this section.

VHH in IDS as a Function of Infant Age

Based on the H&H theory, it is reasonable to hypothesize that VHH in IDS would vary with infant age. Less hyperarticulation is expected at early ages, when the parent probably does not expect the child to understand what he or she is saying but rather speaks to convey affection and maintain a social/emotional connection, while more hyperarticulation is expected as the child starts to understand more of the linguistic content. However, studies explicitly investigating VHH in IDS across development generally report no differences between ages. This is the case regardless of whether they find overall hyperarticulation in IDS compared to in ADS (Liu et al., 2009; Cristia and Seidl, 2014; Wieland et al., 2015; Hartman et al., 2017; Kalashnikova and Burnham, 2018; Kalashnikova et al., 2018), overall hyperarticulation in IDS compared to in ADS (Englund and Behne, 2006; Benders, 2013) or no difference in VHH between IDS and ADS (Xu Rattanasone et al., 2013; Burnham et al., 2015; Wieland et al., 2015).

To date, only a small number of studies have found a relationship between VHH and infant age. Japanese mothers reading to their 6- to 22-month-old infants showed decreasing degree of hyperarticulation with increasing infant age (Dodane and Al-Tamimi, 2007). In Cantonese, age-related differences implied the opposite pattern, with IDS to 3-month-olds being hypoarticulated compared to ADS, whereas no difference in VHH was found between ADS and IDS to 6-, 9- and 12-month-olds (Xu Rattanasone et al., 2013).

Collating results reported in the literature (see Figure 1) also supports the notion that there are no clear systematic differences in VHH related to the age of the infant. Hyperarticulated IDS has been reported for ages between 3.1 months (Kuhl et al., 1997) and 63 months (Liu et al., 2009; not within range in Figure 1), and hypoarticulated IDS has been reported for ages between 0.9 months (Englund and Behne, 2006) and 15.3 months (Benders, 2013). Ages for which no difference in VHH between IDS and ADS has been found span 3.0 months to 20.4 months (Dodane and Al-Tamimi, 2007; Kondaurova et al., 2012; Xu Rattanasone et al., 2013; Burnham et al., 2015; Wieland et al., 2015; Kalashnikova and Burnham, 2018).

To summarize, there appears to be no systematic relationship between VHH in IDS and infant age. Most studies that explicitly investigate VHH in IDS across development report no effect of infant age, and studies that do report age-related differences diverge on whether IDS is more or less hyperarticulated with increasing infant age. Across studies, hyperarticulated IDS, hypoarticulated IDS and no difference between speech registers is reported over very overlapping age spans. From an H&H point of view, this apparent lack of relationship between VHH and infant age is unexpected, but can possibly be attributed to a measurement issue. Vowel space area is a very blunt measure, unable to capture the highly variable dynamics of articulatory

Frontiers in Communication | www.frontiersin.org 2 October 2020 | Volume 5 | Article 523768
Vowel Hyperarticulation in Swedish IDS

FIGURE 1 | VHH in IDS vs. in ADS by infant age. X-axis: Infant age in months. Listed above the x-axis are studies reporting ages at which IDS has been found to be more hyperarticulated than in ADS, below the x-axis are listed studies reporting ages at which IDS has been found to be more hypoarticulated than ADS. Note that the length of the line does not indicate strength of hypo- or hyperarticulation. Dots on the x-axis show ages at which no difference in VHH between IDS and ADS has been reported (Dodane and Al-Tamimi, 2007; Kondaurova et al., 2012; Xu Rattanasone et al., 2013; Burnham et al., 2015; Wieland et al., 2015; Kalashnikova and Burnham, 2018).

VHH in IDS as a Function of Language

Since speaker adaptation is considered a universal trait, similar patterns of VHH in IDS across languages would have been expected according to the H&H theory. Interestingly however, based on the studies published to date, there appears to be less overlap in VHH findings between different languages than between different infant ages (Table 1).

IDS has been reported to be hyperarticulated compared to ADS in Hungarian (Gergely et al., 2017), Mandarin (Liu et al., 2003, 2009; Tang et al., 2017), Russian (Kuhl et al., 1997), and Swedish (Kuhl et al., 1997). Both hyperarticulated IDS and no difference between speech registers have been reported for American English (Kuhl et al., 1997; Kondaurova et al., 2012; hyper: Cristia and Seidl, 2014; Wieland et al., 2015; no difference: Kondaurova et al., 2012; Burnham et al., 2015; Hartman et al., 2017), Australian English (hyper: Burnham et al., 2002; Lam and Kitamura, 2012; no difference: Cristia and Seidl, 2014; hyper: Gergely et al., 2017; Kondaurova et al., 2012; Burnham et al., 2015; no difference: Kalashnikova and Burnham, 2018, 2018, 2018; British English (hyper: Uther et al., 2007, no difference: Dodane and Al-Tamimi, 2007) and Japanese (hyper: Andruski et al., 1999; Miyazawa et al., 2017, no difference: Dodane and Al-Tamimi, 2007). In French, no difference in VHH between registers was found (Dodane and Al-Tamimi, 2007). In Cantonese, no difference was found between ADS and IDS to infants at 6, 9 or 12 months, but IDS to 3-month-olds was hypoarticulated compared to ADS (Xu Rattanasone et al., 2013). For Dutch (Benders, 2013) and Norwegian (Englund and Behne, 2006), hypoarticulated IDS compared to ADS has been reported.

In sum, for any given language in which a difference in VHH has been found between IDS and ADS, either hyperarticulated or hypoarticulated IDS has been reported, never both. This gives the impression that realization of VHH in IDS relative to ADS is language-specific. It is important to note however, that for most languages other than English, results stem from a single group of parent-infant dyads (Russian, Swedish: Kuhl et al., 1997; Norwegian: Englund and Behne, 2006; French:
TABLE 1 | Languages for which hyperarticulated IDS, hypoarticulated IDS, or no difference between speech registers has been reported.

| Hyperarticulation (n = 8) | No difference or hyperarticulation | No difference only | Hypoarticulation only |
|--------------------------|-----------------------------------|--------------------|-----------------------|
| Hungarian                | Am. English                       | French             | Cantonese             |
| Mandarin                | Au. English                       |                    | Dutch                 |
| Russian                  | Br. English                       |                    |                       |
| Swedish                  | Japanese                          |                    |                       |

N refers to number of languages.

In a number of other studies, the vowels are sampled from spontaneous interaction rather than from intentionally elicited target words. The situational context in which the vowels are uttered does not differ between the IDS and ADS samples as clearly as in the studies above, that is, also vowels from words that are not necessarily expected to be strongly emphasized in only one sample are selected for study. Indeed, the studies in which situational context is more similar across samples in this sense report more mixed results. The majority report hypoarticulation in IDS compared to in ADS (Englund and Behne, 2006) or mixed results, with no difference between IDS and ADS for one out of two studies (Wieland et al., 2015) or for one out of two vowel space area measures (Kondaurova et al., 2012), but two report hyperarticulated IDS only (Kuhl et al., 1997; Miyazawa et al., 2017).

To summarize, neither vowel peripherality nor situational context has been explicitly studied in terms of their impact on VHH in IDS, but the body of literature suggests that the situational context may have some influence on the comparison between speech samples, at least for peripheral vowels (the ones used to calculate vowel space area). The fact that both vowel peripherality and situational context are expected to impact realization of VHH, regardless of speech register, highlights the importance of taking methodological choices into account both when interpreting the results and making comparisons across studies.

VHH in IDS by Other Measures

Studies using measures other than vowel space area also report diverging findings on VHH in IDS compared to in ADS. One commonly used way to assess differences in VHH is to compare formant values directly. Importantly, differences in formant frequencies indicate different things depending on which vowel is considered. For example, a lower $F_2$ is indicative of hyperarticulation in /a/ but hypoarticulation in /i/, whereas a lower $F_1$ is indicative of hyperarticulation for both of those vowels. This means that VHH is reported for formants and vowels separately. Therefore, findings of VHH in IDS using this measure will here be presented by type (hypo- or hyperarticulation) and formant, for different vowels.

The high degree of vocal affect typically found in IDS (Kitamura and Burnham, 2003) can result in overall higher formant frequencies (Benders, 2013; Kalashnikova et al., 2017). This means that a finding of higher $F_1$ in /a/ or /o/ in IDS...
compared to in ADS, may not in fact be indicative of vowel hyperarticulation, especially if \( F_1 \) is also higher in other vowels. To ensure that findings are indicative of VHH rather than high vocal affect, only results where formants are lower in IDS than in ADS are considered here.

Hyperarticulation of \( F_3 \) has been reported for /\( i / \) (Englund, 2018), /\( i / \) and /\( u / \) (Kuhl et al., 1997), as well as /\( i / \) and /\( ou / \) (McMurray et al., 2013), and hyperarticulation of \( F_2 \) has been found only for /\( iu / \) (Andruski and Kuhl, 1996; Kuhl et al., 1997; Kondaurova et al., 2012; Burnham et al., 2015; Wieland et al., 2015). Hypoarticulation has been reported only for \( F_2 \) in /\( i / \) (Dodane and Al-Tamimi, 2007). It is important to keep in mind that in all studies reporting on formant frequency difference, several of the individual comparisons revealed no difference between IDS and ADS (Andruski and Kuhl, 1996; Kuhl et al., 1997; Davis and Lindblom, 2000; Englund and Behne, 2005; Dodane and Al-Tamimi, 2007; Green et al., 2010; Kondaurova et al., 2012; McMurray et al., 2013; Wieland et al., 2015; Tang et al., 2017; Englund, 2018).

When it comes to other measures of VHH, a study reporting vowel peripherality (based on acoustic measures) reported hyperarticulated IDS compared to ADS (Wang et al., 2015). On the other hand, a study looking at an articulatory measure of VHH, specifically the position of a point mid-dorsally on the tongue in the sagittal plane of the vocal tract, report no difference between IDS and ADS (Kalashnikova et al., 2017), nor does a study comparing number of reduced instances of vowels based on phonetic transcriptions (Lahey and Ernestus, 2014).

To summarize, no systematic pattern of VHH in IDS can be seen when using other types of VHH measures than vowel space area. Evidence for both hyperarticulation and hypoarticulation of vowels in IDS compared to ADS has been reported, as well as no difference between speech registers. However, most measures other than vowel space area and formant frequency comparison have been used in a single study only, so conclusions should be drawn with caution.

Methodological Choices in Studies of VHH in IDS

Studies on VHH in IDS differ in terms of methodological choices regarding speech sample, vowel token sample and operationalization of VHH. All of those choices impact the dynamics of VHH beyond any comparisons between speech registers.

When it comes to speech samples, the majority of the studies use conversational speech for both the IDS and the ADS condition. Since VHH is believed to be a result of the speaker's in-the-moment adaptation to the needs of the listener, this is a valid setting for comparisons between speech registers, provided all other things are kept constant. A few studies have looked at read speech or similar (Dodane and Al-Tamimi, 2007; McMurray et al., 2013; Burnham et al., 2015; Gergely et al., 2017). This is worth noting, in particular in regards to the ADS condition in which no listener was necessarily present to read the text to. The processes behind VHH variability in “laboratory speech”—when participants are asked to read word lists or texts into a microphone—are not necessarily the same as those hypothesized to underlie VHH variability in conversational speech, where the speaker's assumptions about the listener impacts the clarity of his or her speech. When reading into a microphone, with no listener specified as addressee, it is possible that VHH dynamics in this case is driven primarily by phonetic and linguistic factors such as segmental context, stress and speech rate. This puts into question whether comparisons between IDS and ADS samples are relevant, as they potentially tap into different processes. It is of course possible that the speaker imagines a listener, or considers the experimenter/researcher to be the listener, but nevertheless the online adaptiveness characteristic to conversational speech is disrupted, and generalizability to conversational speech is not given.

Vowel token samples differs between studies, which is likely to impact the results. Most studies have looked at stressed syllables in a small number of target content words (e.g., Andruski et al., 1999; Xu Rattanasone et al., 2013), and only one has included all point vowel tokens regardless of stress and word type (Englund and Behne, 2006). Importantly, due to the dynamic nature of speech production, VHH can vary even within a single utterance, with a more hyperarticulated vowel in the stressed syllable of the focal word than in the unstressed syllables and words. Considering that there is a limit to how hypoarticulated vowels in unstressed positions can be, it is reasonable to assume that the VHH of unstressed vowels is more similar across speech registers than the VHH of stressed vowels. This means that studies with different sampling methods are not directly comparable.

Further on the topic of vowel inclusion criteria, it is important to consider limitations when it comes to formant estimation. Reliability of formant estimations decreases drastically with increasing \( f_o \), both when using manual and automatic formant tracking methods (Lindblom, 1962; Monsen and Engelbreton, 1983). This is not due to limitations in the person or software performing the tracking, but purely a result of the information just not being present in the signal. This issue is of course of vital importance when studying IDS in particular, since a key characteristic of IDS is its high and variable \( f_o \) (Fernald et al., 1989). Yet with few exceptions (Burnham et al., 2015; Wieland et al., 2015), studies of VHH in IDS typically do not report excluding vowels based on high \( f_o \). Not applying a \( f_o \)-based exclusion criterion raises concerns about how reliable the reported findings of VHH actually are.

Lastly, the operationalization of VHH, that is, the specific measure used, can also impact the findings and conclusions. In particular, operationalizing VHH as relative vowel space area heightens the chances of finding VHH differences between samples, as it is calculated based on peripheral vowels, which are most susceptible to variations in VHH. The same is of course true for direct comparisons of formant values of point vowels only, but the difference is that in vowel space area it is more or less inherent to the measure itself, whereas when comparing formant values, it is an issue of sample selection. Using vowel space area as a measure of VHH, one should therefore be aware that the measure most likely overestimates hyperarticulation, and findings should not be generalized to all vowels or the conversation overall.
The Present Study

In the present study, we investigate VHH in Swedish IDS to 12-month-olds, operationalizing VHH in a distance measure that normalizes across speakers and vowels. The purposes of the study are:

1. To determine whether Swedish IDS to 12-month-olds differ from Swedish ADS in terms of VHH, either in terms of degree and frequency.
2. To introduce a novel measure, the *vhh-index*, which operationalizes VHH in a phonetically sound way while permitting analysis of fine-grained VHH dynamics.

VHH in Swedish has previously been compared between ADS and IDS to 3-month-olds, measured with both vowel space area and mean formant frequency comparison. Both measures reveal vowels to be hyperarticulated in IDS relative to ADS (Kuhl et al., 1997). Two previous studies have investigated VHH in IDS to 12-month-olds, one reporting hyperarticulated IDS relative to ADS (Mandarin; Tang et al., 2017) and one finding no difference between speech registers (Cantonese; Xu Rattanasone et al., 2013).

We operationalize VHH using the *vhh-index*, which makes it possible to determine degree of VHH of individual vowel tokens on a continuous scale of hypo- and hyperarticulation, determined by speaker- and vowel-specific characteristics. For each speaker, the mean formant values for all vowel tokens across types and speech registers, define the hypoarticulation endpoint of the scales in acoustic space. This is the centroid of the vowel space for a given speaker, and can be conceptualized as the theoretical vowel they would produce with completely relaxed articulators. Next, the average formants of each vowel type are calculated for each speaker, across speech registers. These determine the midpoints of the vowel type scales, as they represent an individual theoretical “typical” articulation of each vowel. For ease of conceptualization, the hyperarticulation “endpoint” is placed at an equal distance from the scale midpoint as the hypoarticulation endpoint is, but in the opposite direction in acoustic space (Figure 2, left). Importantly, the hyperarticulation endpoint of the scale does not represent a theoretical maximum of hyperarticulation, and individual vowel tokens can thus be more hyperarticulated than this conceptual endpoint. Once individual scales of VHH have been calculated for each speaker and vowel, individual vowel tokens can be placed along these scales. The Euclidian distance in acoustic space between a vowel token and the hypoarticulation midpoint over the distance between the hypo- and hyperarticulation endpoints is the token’s *vhh-index*. This operationalization accounts both for variation in vocal tract size across speakers as well as for differences in absolute formant values across vowel types, making the *vhh-index* of vowel tokens directly comparable across speakers and vowel types. A higher *vhh-index* indicates more hyperarticulation, a lower *vhh-index* indicates more hypoarticulation.

In the present study, we investigate the overall difference in *vhh-index* between IDS and ADS, and compare the results with differences in vowel space area and mean formant values. However, the *vhh-index* being calculated on a token basis allows for more nuanced analysis of the dynamics of VHH within the two speech samples. For example, an overall hyperarticulation of IDS compared to ADS can be a result either of all vowel tokens being somewhat more hyperarticulated in IDS than in ADS, or a smaller number of tokens being extremely hyperarticulated while the rest are relatively similar between samples. To assess the differences in dynamics of VHH, we use several metrics in addition to the mean value to compare IDS and ADS (see Figure 3); range, modal value and peak ratio. Range assesses the difference in overall scope of VHH. The modal value assesses whether IDS tokens differ from ADS tokens overall. Peak ratio (percentage of tokens that have higher *vhh-index* than the 75th percentile, as calculated based on both samples combined) assesses whether there are more highly hyperarticulated tokens in IDS than in ADS.
Our IDS speech sample consists of free play sessions between infants and their parents, during which a small part of the session was dedicated to playing with toys named to elicit point vowels. All vowel tokens, regardless of vowel type, word type and stress, are considered for analysis. Central standard Swedish has nine vowels, most of them with a "long" and a "short" version, but typically the pairs also differ in spectral quality. The front vowels are /i:/, /æ/, /ɛ:/, and /ɛ/; and the front rounded vowels /y:/, /œ/, /œ:/, /o:/, and /u:/ with the more centered variant /a:/.

Our primary focus is the speaker’s behavior rather than the language environment of the infant, since we assume that the bases for VHH in IDS are the same as in spontaneous conversation between adults, that is, the speaker’s in-the-moment adaptation to ensure successful communication based on the perceived or assumed needs of the listener. We therefore use an acoustic frequency scale.

**METHODS AND MATERIALS**

**Participants**

The participants were originally 26 parents (18 mothers, 8 fathers), recorded when interacting in Swedish with their infants and with an adult experimenter. Seven participants were removed after data processing but prior to analysis due to too few vowel tokens ($n < 75$) in either IDS or ADS, resulting in 19 included participants (12 mothers, seven fathers). Parents participated in a longitudinal interaction study at Stockholm Babylab, in which they visited the laboratory every 3–6 months when the child was between 3 months and 3 years. Two of the recorded sessions were used in the present study. The IDS speech sample was obtained from the recording from when the infant was ~12 months, and the ADS sample was obtained from the recording when the infant was ~27 months old. The experimenters in the ADS-sessions were two females and one male and all parents except two had met their experimenter before, up to five times, in previous recording sessions. All parents were native speakers of Swedish, one with an additional native language. All parents had completed high school and many of them ($n = 15$) had a University education.

Mean age of infants (nine girls and 11 boys including a pair of twins) was 12.0 months (range = 11.5–12.3, $sd = 0.2$). All infants were full-term (born within 3 weeks of due date). One child appeared somewhat delayed in cognitive development (according to parent report, no diagnoses), and one child had ear- and hearing-related problems during his first year of life (according to parent report). The rest reported no major/chronic health problems. All infants were monolingual, defined as both parents speaking only Swedish with the child.

**Recordings**

The recording sessions took place in a comfortable carpeted studio equipped with a number of age-appropriate toys in the Stockholm Babylab at the Phonetics laboratory at Stockholm University. All sessions were recorded on video and audio. Four cameras were used to capture all angles of the parent interacting with the child, three of them were mounted on the walls (Canon XA10) and one camera on the chest of the parent (GoPro Hero3). Three microphones were used to get high quality audio recordings, one room microphone (AKG SE 300 B) and two omni-directional wireless lavalier microphones (Sennheiser EW 100 G2) attached to the upper chest of the parent and the child (on a small vest worn by the child), respectively. In the current study, only the sound from the parent lavalier microphone were used in order to get a high-quality close-up channel of the adult’s speech with minimal interference from the child’s vocalizations.

In the IDS condition the parent was instructed to play with the child as if at home. Three toys were present, named to elicit the point vowels (/li:/, /nu:/, and /nu:/). After ~10 min the experimenter entered the room and played a game with the child. In the ADS condition the experimenter had a conversation with the parent at the beginning of the session. The experimenter asked about how they felt about their participation in the study and other open questions about the child to encourage the parent to speak as much and freely as possible for about 2–5 min.

---

1 The longitudinal study was part of the MINT-project, PI Tove Gerholm, funded by Marcus and Amalia Wallenberg Foundation (MAW 2011.007).
**Markup**

As a rule, the recorded sessions were annotated using (ELAN, 2013-2018); Sloetjes and Wittenburg (2008), by a team of researchers and research assistants according to a comprehensive protocol, including both transcription of vocal behavior and annotation of non-vocal behavior (Gerholm, 2018). In this study however, only the transcription of the parents’ speech was used. Therefore, on a number of files, the full annotation protocol was not applied, only the transcription of the parents’ speech was done. In the transcription tier, onset and offset of utterances were marked, and the utterance was quasi-orthographically transcribed. Special types of voice qualities (e.g., whispering, singing, cartoony voice), as well as some meta-information (e.g., if an utterance was an imitation, repetitions, interrupted words) were marked up. Since automated formant estimations are based on assumptions that are valid for regular voice quality, all utterances in special voice qualities were excluded from further analysis. All markup regarding meta-information was also removed for this study. The cleaning of the transcription was done using a script written in R 3.5.0 (R Core Team, 2018). In both conditions we used speech only from the parent, and only parent speech directed to the child in the IDS condition and parent speech directed to the experimenter in the ADS condition.

**Acoustic Measures**

Formants were estimated using Praat 6.0.43 (Boersma and Weenink, 2018) with default settings (time step set to 0 s, analysis window length set to 0.025 s, pre-emphasis set from 50 Hz) except maximum formant ceiling. Maximum number of formants was set to between 4 and 6.5, depending on the formant ceiling setting (in accordance with Praat’s recommended default). To ensure robust formant estimations, the formant ceiling was optimized for each speaker and vowel type (Escudero et al., 2009). For each recording, formants were estimated with the maximum formant ceiling setting ranging from 4,000 to 6,500 Hz in steps of 10 Hz resulting in 251 formant estimation outputs, after which the formant ceiling resulting in the smallest variance for each vowel type was selected (separately for F1 and F2). The formant ceiling selection procedure was implemented in a script written in R 3.5.0 (R Core Team, 2018). Formants were estimated for the entire recordings, for later extraction F1 and F2 values for individual vowel tokens (see Data processing).

The \( f_o \) was estimated using Praat 6.0.37 (Boersma and Weenink, 2018) in order to be able to exclude vowel tokens based on \( f_o \). The \( f_o \) estimation was carried out in two steps to minimize potential pitch tracking errors. The first step was to generate a pitch object using Praat’s pitch extraction (autocorrelation) with the default settings and the floor set to 60 Hz and ceiling to 600 Hz. In the second step, this pitch object was used to define the pitch floor and the pitch ceiling in another pitch extraction. The floor was set to \( q_15 \times 0.83 \) and the ceiling was set to \( q_65 \times 1.92 \) (De Loose and Hirst, 2010), where \( q_15 \) and \( q_65 \) are the 15th and 65th percentile of the original pitch object. The \( f_o \) for the entire file was estimated, for later extraction of values at the time of individual vowel tokens (see Data processing).

**Data Processing**

The cleaned quasi-orthographic transcriptions were automatically segmented and converted to phonemic SAMPA transcriptions using the web service Chunk Preparation 2.32 and 2.33 of the Bavarian Archive for Speech Signals at the University of Munich (Reichel and Kisler, 2014). The settings used were language: Swedish, sampling rate: 16,000, input format: eaf, input tier name: ORT and keep annotation: no. The phonemic SAMPA transcriptions were then converted to IPA and aligned with their audio files using the web service WebMAUS General 5.33 of the Bavarian Archive for Speech Signals at the University of Munich (Schiel, 1999; Kisler et al., 2017). The default settings were used, except language: Swedish, output format: csv, chunk segmentation: true, output symbols: ipa, MAUS modus: align and relax min duration: true. This automatized procedure results in a phonological transcription on segment level.

A script was written in R 3.5.0 (R Core Team, 2018) to extract the following information for each vowel token: transcription, register, mean \( F_1 \), mean \( F_2 \), median \( f_o \), duration, immediately preceding segment (if any), and immediately following segment (if any). Mean formant values were calculated based on the center 40% of the vowel token’s duration. Median \( f_o \) was used rather than mean, since it is more robust to the type of errors sometimes encountered in automatic \( f_o \) estimation. Vowel tokens with an \( f_o \) of 350 Hz or higher were excluded from further processing since formant estimations performed on them would not be reliable (Monsen and Engebretson, 1983). Segmental context was coded for place of articulation (bilabial, dental or velar context) and nasality (e.g., if either the preceding or the following segment was a nasal, context nasality was set to 1, otherwise to 0), since these are attributes of the segmental context that can strongly impact the formants of the vowel (Davis and Lindblom, 2000).

**Calculating the vhh-index**

For each participant and vowel type, the mean \( F_1 \) and mean \( F_2 \) was calculated, and set as the mid-point of the VHH scales (one per vowel type). The centroid, that is, the mean of all vowel types, was designated the hypoarticulation endpoint of the scale. For ease of conceptualization, hyper-articulation endpoints of the VHH scales (one per vowel type) was calculated as twice the distance between the centroid and the mid-point (Figure 2, left).

For each vowel token, the distance (in \( F_1-F_2 \) vowel space) to the centroid was calculated, and divided by the distance to the hyperarticulation endpoint. This results in a percentage, normalized for vowel type and speaker, that is a measure of

| TABLE 2 | Number of tokens excluded due to too high \( f_o \), due to too few tokens per participant, or due to being outliers, as well as number of included tokens. |
|---------|---------------------------------------------------------------------------------------------------------------|
|         | Unable to estimate formants | Unable to estimate \( f_o \) | Too high \( f_o \) | Outliers | Too few tokens | Included tokens |
| ADS     | 136                         | 563                          | 137                | 282      | 277            | 4,496           |
| IDS     | 292                         | 2,034                        | 1,005              | 1,262    | 2,329          | 7,688           |

“Too few tokens” refers to the number of tokens excluded as a result of excluding seven participants due to having <.75 tokens total in either IDS or ADS.
the degree of vowel hypo- or hyperarticulation and referred to as the vhh-index of the token (Figure 2, right). Higher vhh-index means more hyperarticulation, lower vhh-index means more hypoarticulation. In order to be compatible with previous studies, vowel space area (calculated as described in e.g., Liu et al., 2003) and mean formants were also compared between speech registers.

**VHH Dynamics**

To capture potential differences in the dynamics of VHH in ADS vs. IDS, the following three metrics were used: range, mode, and peak ratio. Together, these metrics paint a more nuanced picture of how VHH varies within a speech sample than a comparison of vowel space area (see Figure 3).

**RESULTS**

**Robust Formant Estimation**

Three steps were taken to ensure robust automatic formant estimation: (1) excluding all vowel tokens with a median $f_o$ of 350 Hz or more, (2) using optimal formant ceiling setting per speaker and vowel type, and (3) excluding all tokens in which either $F_1$ or $F_2$ differed more than two standard deviations from the mean formant value (calculated separately for speaker sex and vowel types, but pooled across registers). Although it would have been ideal to check the estimated formants against expected ranges as part of the process to ensure robust formant estimation, it was not possible in this case since there is a lack of reference data on Swedish short vowels in the literature. Table 2 provides information about number of included and excluded tokens at each stage. Figure 4 shows point vowel tokens remaining after exclusion based on $f_o$ with formants estimated using a generic formant ceiling setting based on speaker sex, the same remaining

| Context  | df | t   | p   |
|----------|----|-----|-----|
| Bilabials| 18 | 0.277 | 0.785 |
| Dentals  | 18 | 1.18 | 0.253 |
| Nasals   | 18 | 0.829 | 0.418 |
| Velars   | 18 | −0.226 | 0.824 |

Bonferroni-corrected alpha is 0.013.

**Table 2** | Paired-samples $t$-test comparing the proportion of vowel tokens with different segmental contexts in IDS and ADS.

![Figure 4](Frontiers in Communication | www.frontiersin.org) 9 October 2020 | Volume 5 | Article 523768

**Figure 4** | Illustration of formant estimations. Top: female speakers. Bottom: male speakers. Left: generic formant ceiling. Right: optimal formant ceiling. Note that outliers (in terms of either $F_1$, $F_2$ or both, estimated using optimal formant ceiling settings) are not included in the illustration.
vowel tokens but estimated using individually adapted formant ceiling settings (Escudero et al., 2009).

**Phonetic Context**

Since the phonetic context contributes to the variation in VHH, we compare our IDS sample and our ADS sample on this point. T-tests were performed to compare the proportion of vowel tokens with different contexts in IDS and ADS (Table 3, see also Figure 5). No significant differences were found, suggesting that the distribution of phonetic contexts was similar for vowel tokens in ADS and in IDS.

Mean and median vowel duration, as well as standard deviation in ADS and IDS are presented in Table 4, and the distributions can be found in Figure 6. A t-test revealed that vowel duration differed significantly between ADS and IDS \(t(18) = -2.47, p = 0.024\).

**Comparison With Other Measures**

A paired-samples t-test showed that the vowel space area differed significantly between ADS and IDS \(t(18) = 4.35, p < 0.001\), see Figure 7 (left). Differences in formant values between speech registers were assessed using six t-tests, one per formant and point vowel (Figure 7 right, Table 5). The numerical difference in mean formant value in most cases indicate vowel hyperarticulation in IDS compared to in ADS, but only \(F_2\) of /i/ was significantly hyperarticulated, and \(F_2\) of /al/ was significantly hypoarticulated.

For comparison with vowel space area, the average vhh-index for all point vowels combined (per participant) was used, and for comparisons with formant measures, the average vhh-index for each individual point vowel (per participant) was used. T-tests were performed to compare the mean vhh-index between IDS and ADS (Table 6). In most cases, the mean vhh-index was higher for IDS than for ADS and thus indicative of more hyperarticulation in IDS than in ADS, but the difference was only significant for /i/.

To make use of one of the benefits of the vhh-index over vowel space area, namely the fact that the degree of VHH can be estimated for each individual token, the effect of register on VHH was also analyzed using linear mixed effect models, using the lmer package (Bates et al., 2015) in R. The explained variable was vhh-index, and the factor of interest was register (IDS, ADS). Duration and vowel type (/i/, /I/, /e/, /e/, /æ/, /æ/, /a/, /a/, /u/, /a/) were also included as fixed factors, since they are both expected to impact VHH. Random slope
effect was participant. To test the impact of register, an identical model was created except that register was not included as a fixed effect, and a likelihood ratio test was performed. This revealed a significant effect of register on vhh-index [$\chi^2(1) = 23.4, p < 0.001$].

### Measures of Dynamics

For each of the three measures of dynamics (range, mode, peak ratio), paired-samples $t$-test was performed (see Table 7). The results show that when all vowels were considered, the range of VHH was not necessarily larger in IDS than in ADS, but both modal value and peak ratio were greater in IDS than in ADS, indicative of hyperarticulation in IDS. When point vowels were excluded to account for potential hyperarticulation due to emphasis on target words (toy names), the same pattern was found.

### DISCUSSION

Investigating VHH in Swedish IDS to 12-month-olds, the present study finds hyperarticulation relative to ADS. This difference was found using both the traditional measure of vowel space area and using the vhh-index metric in a token-based analysis including all vowels. This is in line both with previous reports on Swedish IDS (to 3-month-olds; Kuhl et al., 1997) and with previous findings in IDS to 12-month-olds in various other languages (Liu et al., 2003; Kondaurova et al., 2012; Cristia and Seidl, 2014; Kalashnikova et al., 2017, 2018; Tang et al., 2017; Kalashnikova and Burnham, 2018). However, in IDS to 12-month-olds, there has also been reports of hypoarticulated IDS (Benders, 2013) or no difference between registers (Dodane and Al-Tamimi, 2007; Kondaurova et al., 2012; Xu Rattanasone et al., 2013; Burnham et al., 2015).

The study presents comparisons between the vhh-index and two traditional measures, vowel space area and mean formant frequencies. Since vowel space area is based on point vowels and calculated on a subject level, the vhh-index for all point vowels combined was calculated for each subject and compared across registers. The vowel space area measure resulted in a significant difference between IDS and ADS, but the vhh-index analysis did not. This is likely due to the fact that differences in each formant contributes separately to the vowel space area difference, whereas in the vhh-index both formants are aggregated into a single estimation of VHH. Taking the vowel /a/ in Figure 7 (left)

![Figure 7](image_url)
In conclusion, the present study demonstrates that vowels in Swedish IDS to 12-month-olds are hyperarticulated compared to vowels in ADS, both in terms of degree and frequency of occurrence. The study also introduces a novel way to quantify VHH, the vhh-index, which permits assessing dynamics of VHH within a sample or conversation, collapsing analyses across speakers and vowels, and performing analysis on the token level.

**DATA AVAILABILITY STATEMENT**

The dataset for this study cannot be publicly shared since the consent obtained from participants of the original project does not include public sharing of material derived from their participation.
ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Regional Ethics Committee in Stockholm (2015/63-31). Written informed consent to participate in this study was provided by the participants’ legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

EM: study conceptualization, drafting of the manuscript, and analysis and interpretation. EM and LG: study design, data processing, and critical revisions of the manuscript. LG: data collection (part). All authors contributed to the article and approved the submitted version.

FUNDING

The research presented in this article was part of the HELD-project (PI LG), funded by Riksbankens Jubileumsfond (RJ P17-0175). The material was recorded and partly annotated as part of the MINT-project (PI Tove Gerholm), funded by Marcus and Amalia Wallenberg Foundation (MAW 2011.0070).

ACKNOWLEDGMENTS

We would like to thank all families participating in the MINT-project and Tove Gerholm for sharing the recorded and annotated material with us. We would also like to thank Tove Gerholm, David Pagmar och Ulrika Marklund for taking part in data collection, Freya Eriksson, Alice Gustavsson, Mika Matthis, Linea Rask, Johanna Schelhaas and Sofia Tahbaz for transcribing the material used in this study, and Anna Ericsson for performing part of the automated data processing. Thanks also to Björn Lindblom for feedback at an early stage of the study and to Iris-Corinna Schwarz for valuable comments on an earlier version of the manuscript.

REFERENCES

Adams, S. G., Weismer, G., and Kent, R. D. (1993). Speaking rate and speech movement velocity profiles. J. Speech Lang. Hear. Res. 36, 41–54. doi: 10.1044/jslr.3601.41
Andruski, J. E., and Kuhl, P. K. (1996). “The acoustic structure of vowels in mothers’ speech to infants and adults,” in Proceeding of Fourth International Conference on Spoken Language Processing, 1545–1548. doi: 10.1109/ICSLP.1996.607913
Andruski, J. E., Kuhl, P. K., and Hayashi, A. (1999). “The acoustics of vowels in Japanese women’s speech to infants and adults,” in Proceedings of the 14th International Congress on Phonetic Sciences (San Francisco, CA), 2177–2179.
Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. doi: 10.18637/jss.v067.i01
Benders, T. (2013). Mommy is only happy! Dutch mothers’ realisation of speech sounds in infant-directed speech expresses emotion, not didactic intent. Infant Behav. Dev. 36, 847–862. doi: 10.1016/j.infbeh.2013.09.001
Benders, T., Pokharel, S., and Demuth, K. (2019). Hypo-articulation of the four-way voicing contrast in Nepali infant-directed speech. Lang. Learn. Dev. 15, 232–254. doi: 10.1080/15475441.2019.1577139
Boersma, P., and Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0.37 (February 3, 2018) to 6.0.43 (8 September 2018). Available online at: http://www.praat.org/
Burnham, D., Kitamura, C., and Vollmer-Conna, U. (2002). What’s new, pussycat? On talking to babies and animals. Science 296, 1435–1435. doi: 10.1126/science.1069587
Burnham, E. B., Wieland, E. A., Kondaurova, M. V., McAuley, J. D., Bergeson, T. R., and Dilley, L. C. (2015). Phonetic modification of vowel space in storybook speech to infants up to 2 years of age. J. Speech Lang. Hear. Res. 58, 241–253. doi: 10.1044/2015_JSLHR-S-13-0205
Cristia, A., and Seidl, A. (2014). The hyperarticulation hypothesis of infant-directed speech. J. Child Lang. 41, 913–934. doi: 10.1017/S0305000912000669
Davis, B. L., and Lindblom, B. (2000). “Phonetic variability in baby talk and development of vowel categories,” in Chapter in Emerging cognitive abilities in infancy, eds Lacerda, von Hofsten, Heineman (Cambridge: Cambridge University Press), 135–171.
De Looze, C. D., and Hirst, D. (2010). “Integrating changes of register into automatic intonation analysis,” in Proceedings of the Speech Prosody 2010 Conference (Chicago, IL).
Dodane, C., and Al-Tamimi, I. (2007). “An acoustic comparison of vowel systems in adult-directed speech and child-directed speech: evidence from French, English and Japanese,” in Proceedings of the 16th International Congress of Phonetics Sciences (Saarbrücken), 6–10.
ELAN. (2013–2018). Nijmegen: Max Planck Institute for Psycholinguistics [Computer program]. Version 4.6.2 (November 07, 2013) to 5.3 (August 22, 2018). Available online at: https://tlaver.tlabtools/elan/
Elmlinger, S. L., Schwade, J. A., and Goldstein, M. H. (2019). The ecology of prelinguistic vocal learning: parents simplify the structure of their speech in response to babbling. J. Child Lang. 46, 998–1011. doi: 10.1017/S0305000919000291
Englund, K., and Behne, D. (2006). Changes in infant directed speech in the first six months. Infant Child Dev. 15, 139–160. doi: 10.1002/icd.445
Englund, K. T. (2018). Hypoarticulation in infant-directed speech. Appl. Psycholinguist. 39, 67–87. doi: 10.1017/S0142716417000480
Englund, K. T., and Behne, D. M. (2005). Infant directed speech in natural interaction—Norwegian vowel quantity and quality. J. Psycholinguist. Res. 34, 259–280. doi: 10.1007/s10366-005-3640-7
Engrand, S. (1999). “Swedish,” in Handbook of the International Phonetic Association: A Guide to the usage of the International Phonetic Alphabet, (Cambridge: Cambridge University Press), 140–142. Available online at: https://www.internationalphoneticassociation.org/content/handbook-IPA
Escudero, P., Boersma, P., Rauber, A. S., and Bion, R. A. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. J. Acoust. Soc. Am. 126, 1379–1393. doi: 10.1121/1.3180321
Fant, G., and Kruckenberg, A. (1962). “Notes on stress and word accent in Swedish,” in Proceedings of the International Symposium on Prosody (Yokohama), 2–3.
Fernald, A., Papousek, M., de Boysson-Bardies, B., Papousek, H., Do, Y. H., and Fukui, I. (1989). A cross-language study of prosodic modifications in mothers’ and fathers’ speech to preverbal infants. J. Child Lang. 16, 477–501. doi: 10.1017/s0305000900010679
Gergely, A., Faragó, T., Galambos, Á., and Topál, J. (2017). Differential effects of different vocalization types on infants’ vocal imitation. Frontiers in Communication | www.frontiersin.org 13 October 2020 | Volume 5 | Article 523768
Gerholm, T. (2018). Conventions for Annotation and Transcription of the MINT-Project: Modulating Child Language Acquisition Through Parent-Child Interaction, MAW:2011.007. Available online at: http://su.diva-portal.org/smash/get/diva2:1204492/FULLTEXT04.pdf
Gerholm, T. (2018). Conventions for Annotation and Transcription of the MINT-Project: Modulating Child Language Acquisition Through Parent-Child Interaction, MAW:2011.007. Available online at: http://su.diva-portal.org/smash/get/diva2:1204492/FULLTEXT04.pdf

Frontiers in Communication | www.frontiersin.org 13 October 2020 | Volume 5 | Article 523768

Gustavsson, Mika Matthis, Linnea Rask, Johanna Schelhaas and Sofia Tahbaz for transcribing the material used in this study, and Anna Ericsson for performing part of the automated data processing. Thanks also to Björn Lindblom for feedback at an early stage of the study and to Iris-Corinna Schwarz for valuable comments on an earlier version of the manuscript.
Wieland, E. A., Burnham, E. B., Kondaurova, M., Bergeson, T. R., and Dilley, L. C. (2015). Vowel space characteristics of speech directed to children with and without hearing loss. *J. Speech Lang. Hear. Res.* 58, 254–267. doi: 10.1044/2015_JSLHR-S-13-0230

Xu Rattanasone, N., Burnham, D., and Reilly, R. G. (2013). Tone and vowel enhancement in Cantonese infant-directed speech at 3, 6, 9, and 12 months of age. *J. Phonet.* 41, 332-343. doi: 10.1016/j.wocn.2013.06.001

Xu, N., Burnham, D., Kitamura, C., and Vollmer-Conna, U. (2013). Vowel hyperarticulation in parrot-, dog- and infant-directed speech. *Anthrozoös* 26, 373–380. doi: 10.2752/175303713X13697429463592

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Marklund and Gustavsson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.