UvA-DARE (Digital Academic Repository)

Exploring the acoustic vowel space in two-year-old children: results for Dutch and Hungarian

van der Stelt, J.M.; Zajdó, K.; Wempe, T.G.

DOI

[10.1016/j.specom.2005.06.006](https://doi.org/10.1016/j.specom.2005.06.006)

Publication date

2005

Published in

Speech Communication

[Link to publication](#)

Citation for published version (APA):

van der Stelt, J. M., Zajdó, K., & Wempe, T. G. (2005). Exploring the acoustic vowel space in two-year-old children: results for Dutch and Hungarian. *Speech Communication*, 47(1-2), 143-159. <https://doi.org/10.1016/j.specom.2005.06.006>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Speech Communication xxx (2005) xxx–xxx

SPEECH
COMMUNICATIONwww.elsevier.com/locate/specom

Exploring the acoustic vowel space in two-year-old children: Results for Dutch and Hungarian

Jeannette M. van der Stelt ^{a,*}, Krisztina Zajdó ^b, Ton G. Wempe ^a

^a Institute of Phonetic Sciences/ACLC, University of Amsterdam, Herengracht 338, 1016 CG Amsterdam, The Netherlands

^b Division of Communication Disorders, University of Wyoming, Dept. 3311, 8 Ross Hall, Laramie, Wyoming 82071-3311, USA

Received 9 November 2004; received in revised form 23 June 2005; accepted 24 June 2005

Abstract

In the last decade, there has been an increasing interest in exploring patterns of vowel acquisition in young children. Traditionally, researchers attempt to estimate formant values of vowel realizations via acoustic measurements. However, these techniques have yielded questionable results, due primarily to a low sampling rate of the spectrum caused by a high fundamental frequency in young children's speech. Additionally, the researcher's knowledge about the intended vowel quality affects the decision pertaining to vowel formants. A frequency domain band filtering analysis method that minimizes the dependence of the results on F_0 is developed to measure the spectral envelopes in children's utterances automatically, and is applied to existing utterance data sets of Dutch and Hungarian. One further advantage of the current method is that it selects a maximum of 10 measurement points along the length of the utterance. Data reduction of all filter outputs is achieved via Principal Component Analysis (PCA). By using the first 2 eigenvectors, a reference plane is created. The first two eigenvectors account for 54.2 vs. 58.6% in the Dutch and Hungarian data sets, respectively. Next, a common reference plane for Dutch and Hungarian two-year-olds is constructed by balancing the number of utterances that are analyzed per language. Perceptually judged as being correctly pronounced corner vowels of Dutch- and Hungarian-speaking two-year-old boys were mapped onto this common Dutch–Hungarian reference plane. The band filtering method has shown to be robust with regard to signal-to-noise ratios and to the differences in numbers of measurements.

© 2005 Published by Elsevier B.V.

Keywords: Automated band filtering; Vowel acquisition; Principal component analysis; Dutch; Hungarian

* Corresponding author. Tel.: +31 20 5252183; fax: +31 20 5252197.

E-mail address: jeannet.van.der.stelt@uva.nl (J.M. van der Stelt).

1. Introduction

1.1. About vowel acquisition

Investigations into child phonology have traditionally focused on the acquisition of consonants. This observation is especially true for studies examining speech sound development in 2;0 to 5;0 year old children. Previous research (e.g., Stoel-Gammon, 1983; Davis and MacNeilage, 1990) has firmly established that, similarly to speech patterns found in adult speech, consonant production in children is also influenced by consonant–vowel interactions both within and across syllable boundaries (as suggested, for example, in the frame-content theory; MacNeilage, 1998). However, research into vowel acquisition in children acquiring their native language has been fairly neglected.

During development, anatomical and physiological changes take place in the oral cavity. For example, the ratio of tongue mass to the size of the oral cavity decreases, thereby allowing more space for the child's tongue to move around. Further, the pharyngeal area also becomes larger. At the same time, the use of the oral space is also driven perceptually, although it is unclear to what extent. Serkhane et al. (2002) similarly suggest the presence of two mechanisms in speech development: “exploration of the vocal tract sensori-motor abilities, and the imitation (overt simulation) of caretakers' language sounds.” (p. 45). Exploring the vocal tract implies an enormous variability in the quality of vocalic productions in individual children. Analysis of vocalizations from both deaf and hearing infants has already confirmed the presence of a great amount of variability in the individual spectral envelopes (Van der Stelt et al., 2003a).

Most recently, vowel-like vocalizations of young children are coming to the forefront of interest in acoustic research (e.g., Lee et al., 1997; Robb et al., 1997) since the developmental process from infants' vocal production towards adult-like vowels in older children is not yet unraveled. Kuhl and Meltzoff (1996) pointed at vocal imitation as a mechanism that guides the infant along the road from language universal produc-

tion patterns to language-specific vowel productions. Infants separate the imitated adult corner vowels with increasing accuracy as early as 12–20 weeks. Kuhl (1993) argued that six-months-old infants are capable of perceiving and produce sounds that belong to their native-language “in the absence of word acquisition and linguistic contrast” (p. 130). In an earlier paper (Kuhl and Meltzoff, 1982) these capacities were interpreted to “reflect a knowledge of the relationship between audition and articulation” (p. 1140). The work of De Boysson-Bardies et al. (1989) showed that the articulatory patterns for vowel productions of 10-month-old infants already reflect differences in the F_1 – F_2 plane for the ambient language background. Given the smaller vocal tract of young infants, as well as the much higher pitch, these formants confront researchers with the problem of speaker normalization. Kuhl and Meltzoff (1996) discussed this concept of perceptual vowel constancy: in speaker normalization, F_0 and the vowel formants play a fundamental role while learning to speak. Kent (1993) however, preferred to talk about an acoustic discrimination, instead of phonetic perception.

This study is guided by Kent's acoustic point of view, using a band-filtering measurement technique of the spectral envelope (explained below). In the future, we intend to further explore the relations between the articulatory vowel space (F_1 – F_2 plane, which ignores all intensity information) and the acoustical vowel space (using the spectral envelope information).

1.2. Acoustical approaches

Using spectral analysis by LPC inverse filtering to detect vowel formants is not suitable for our purposes to explore the acoustic vowel space, which is explained below. Fig. 1 shows both LPC and bandfilter spectra together with the long-term ‘line’ spectrum of an example of a child's utterance (Dutch subject 4 at age 24 months, one of the /a:/ sounds, with an F_0 of 330 Hz. One F_0 period at time 0.08 s. from the start is isolated and multiples of this single period are concatenated up to a duration of 500 ms. which is needed for calculating a long-term spectrum). The LPC spectra were calcu-

Fig. 1. Comparison of LPC and bandfilter spectra of an /a:/ sound produced by a 24-months-old Dutch boy. LPC spectra calculated with two differently chosen orders (14, the curve in red and 16, the curve blue), bandfilter spectrum using Gauss bandfilter with an effective bandwidth of 360 Hz. The bandfilter envelope is presented by the upper curve (in green). One period of the /a:/ sound was recirculated to achieve a long sound with constant spectral properties. The grey lines represent its longterm spectrum. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

lated with the following characteristics: down sampling to 14 kHz, window length of 25 ms, pre-emphasis of 6 dB/octave from 50 Hz. It can be seen clearly that the bandfilter spectrum follows the line spectrum envelope best, whereas the LPC spectra either do not expose enough peaks (order 14), or identify peaks at the wrong places (order 16). In addition, both LPC graphs display a very narrow peak at the eighth F_0 harmonic. Narrow peaks can lead to high values of intensity variance per frequency “bin” caused by minor shifts in frequency position of such peaks. These shifts cause an undesirable contribution to the variance data.

Spectral analysis by selection of “spectral slices” from the widely used spectrogram that is created by a moving short-term window is not very suitable either for our automated analysis. The consecutive spectra can vary greatly caused by the huge number of “wrong” positions of the window during its moving through the voiced signal (Wempe, 2001 for a technical expose), so that a very high amount of “false” variance can be expected in the spectral component’s intensities.

Because the values at multiples of F_0 in the long-term spectrum are in fact frequency samples of the spectral envelope that is equivalent to the filter function, we applied the bandfilter method, as

this method resembles the spectral envelope best, provided that the filter bandwidth is sufficiently high with respect to F_0 . This automated bandfiltering method was developed as an alternative for formant analysis (Van der Stelt et al., 2003b).

Pols and colleagues have already applied bandfiltering (see Plomp et al., 1967; Pols et al., 1969, 1973; Pols, 1977). Results suggested that the first two principal components strongly correlated with the first two formants in adult vowels that were measured by means of a 1/3-octave bandfilter set. In the current paper we present a different bandfilter method, which can be used to calculate the spectral envelope in infant sounds that are well known for the difficulty in measuring their formants (see Section 3 for a description of this method). The automated method takes care of selecting the measuring positions and ensures that the results are not biased by the perceived sound quality.

Thus, large corpora of both labeled and unlabeled vocalizations can be measured. As Palethorpe et al. (1996) pointed out when classifying vowels in 4-year-old children and adults, one major advantage of using automated analysis over the hand-edited formant measurement method is that the former method results in reliable measurement values regardless of the collected utterance

types. A large-scale comparison over speakers, over age, and over languages is then at least acoustically un-biased.

1.3. About the chosen corpora of Dutch and Hungarian utterances

In earlier work on acoustic analysis of vocalizations of Dutch infants (Van der Stelt et al., 2003a, using an existing corpus of Clement, 2004), the vocalizations examined were linguistically unlabeled and only categorized by means of a sensori-motor system that identifies developmental milestones in the coordination of respiration, phonation and articulatory movements of young babies (Koopmans-van Beinum and Van der Stelt, 1986). This system primarily focuses on the movements of the speech apparatus, representing the consonantal parts in vocalizations, and as such to a lesser extent on the syllable nucleus itself. Yet it is the syllable nucleus that may contain information about developmental changes in the child's vowel production space.

In the present study, vocalizations of monolingual two-year-old boys with a native language of either Dutch or Hungarian are analyzed. The Hungarian bi-syllabic utterances are linguistically labeled, since they are perceptually selected to resemble an adult target word. These target words were analyzed in this study as unlabeled speech data. By comparing existing sound corpora for Dutch and Hungarian children, it can be decided whether the ambient language already influences two-year-old children's vowel productions.

2. Phonetic data sets

2.1. Acoustic data collection in young children

Collecting utterances (longitudinally) from infants and toddlers poses its difficulties. Datasets often result from compromises between the acoustic quality, the spontaneity and/or naturalness of the utterances, and the amount of usable sound tracks. In this study comparing two languages, existing corpora of speech were utilized that differed in various ways. These two languages were selected out of convenience. An attractive aspect is that both languages have vowel systems with a number of similar characteristics, including front unrounded, front rounded, and back rounded series, close height degrees, and similar quantity contrasts (see Figs. 2 and 3).

A difference between the two data sets concerns the recording settings. The Dutch children were recorded in the home environment, while the Hungarian children came to a sound lab. Because of

Fig. 2. The IPA adult vowel inventory of the standard dialect of Hungarian. This dialect lacks diphthongs.

Fig. 3. The IPA adult vowel inventory of the standard dialect of Dutch. Left figure: Dutch monophthongs. Right figure: Dutch diphthongs and diphthongized vowels (upper three).

the possible limitations of the quality of the recordings made at the homes of the subjects, we measured the signal-to-noise ratios of the recorded material automatically using a separate PRAAT script (constructed by Boersma and Weenink, 1996, with regular program updates via a website) which estimated the minimum and peak levels of all intensity contours per month. We assumed that the recording level during the 10-min recordings remained unchanged. For this purpose, we pre-filtered the signals prior to the intensity measurements so that only our frequency range of interest (90–7000 Hz) was taken into account. Table 1, Section 2.2 shows the result for the Dutch, and Table 2, Section 2.3 for the Hungarian recordings. The signal-to-noise ratios (the lowest being 42.3 dB, see Table 1) were judged to be satisfactory for our spectral bandpass analysis (see Section 4.2 Fig. 14 for an investigation of noise dependency on final analyses output).

Further, in the selection module of the analysis script we omit the low intensity fragments of the vocalizations to exclude as much as possible consonantal parts in the utterances (see Fig. 5 for the flow chart of the selection and analysis module).

The majority of utterances were treated as unlabeled data. However, the way of eliciting the utterances differed for the two languages. The Dutch data resulted from spontaneous (uncontrolled)

mother–child interaction, whereas the Hungarian set can be regarded as citation speech. Citation utterances may reflect less reduction in the F_1 – F_2 plane than spontaneous speech. These differences need to be considered when formulating interpretations of the results.

2.2. The Dutch data set

Participants included 5 healthy boys who were included as control subjects in a longitudinal research project examining vocal development in deaf infants (Clement, 2004). All participants were acquiring Dutch as their native language. Parents were instructed to obtain audio recordings of at least 30 min mother–infant interaction in naturalistic home environments monthly, starting as soon as possible after birth up to 18 months of age.

An additional recording was made by the researcher when the children were 24 months old, by means of a handycam Sony recorder type CCD-TR105 and an external microphone AKG-D190C with a sensitivity of 1.6 mV/Pa. Only this recording was considered in the present study. The sounds of both mother and child were recorded in an analogue manner onto the audio tracks of the videotape. The tapes were digitized (48 kHz sample frequency, 16 bit, mono) by a Silicon Graphics computer (Iris Indigo R 4000) afterwards. These recordings were used in the present study. For the minimal and maximal intensities, as well as the signal-to-noise ratios per recorded Dutch child see Table 1. A 10-min segment of ongoing spontaneous vocal mother–infant interaction was selected from each recording of the 5 normally hearing two-year-olds. From that ‘dialogue’, 50 child utterances were selected randomly, categorized by means of the sensori-motor system that identifies developmental milestones in speech motor coordination (e.g. Koopmans-van Beinum

Table 1
Minimal and maximal intensities, and the signal-to-noise ratios in dB in the recordings of the 5 two-year-old Dutch boys

Dutch	Boy 1	Boy 2	Boy 3	Boy 4	Boy 5
Minimum intensity	46.7	34.0	38.1	36.5	18.8
Maximum intensity	89.0	87.9	87.8	84.3	61.3
Signal/noise	42.3	53.9	49.7	47.8	42.5

Table 2
Minimal and maximal intensities, and the signal-to-noise ratios in dB in the recordings of the 8 two-year-old Hungarian boys

Hungarian	Boy 1	Boy 2	Boy 3	Boy 4	Boy 5	Boy 6	Boy 7	Boy 8
Minimum intensity	30.3	30.7	24.9	30.0	29.3	28.7	31.0	29.8
Maximum intensity	88.1	91.0	91.4	91.6	91.8	87.4	93.0	78.8
Signal/noise	57.8	60.3	66.5	61.7	62.5	58.8	62.0	48.9

and Doppen, 2003). The majority of selected vocalizations were produced during egressive respiration with and without interrupted phonation. With regard to articulation, some vocalizations were produced with an unchanged articulatory position; others contained one or more articulatory movements. In this manner, the utterances must be regarded as linguistically unlabeled, since only anatomical and physiological aspects of the sound production process have controlled the selection. This data set contains $5 \times 50 = 250$ utterances of the 5 boys, taken from the recording when they were two years of age.

2.3. The Hungarian data set

Participants included 8 two-year-old boys with uneventful health histories. The children participated in a cross-sectional study on vowel acquisition between the ages of 2;0 and 4;0 years with a six months interval ($n = 112$; Zajdó, 2002a,b; Zajdó and Stoel-Gammon, 2003). Only the data of the two-years-old boys were considered in the present study. All children were acquiring the standard dialect of Hungarian as their native language only. The children were recruited from the Budapest area. Before participation was granted, the mothers' speech was judged to have no pronunciation errors in vowel production (e.g., the presence or absence of distinction between phonologically short and long vowels; individual vowel qualities).

All samples were recorded in a sound-treated room (6 m^2) at the Research Laboratory of Acoustics at the Budapest University of Technology and Economics. Recordings were obtained by using an omni-directional AKG CK 77-3 condenser Lavalier microphone with a sensitivity of 46 mV/Pa . The microphone forwarded signals of both the caregiver and the child to a SONY Vaio PCG-FX190K laptop. The "Sound Forge" acoustic software (Version 5.0, Built 117) was used for digital recording, with recording attributes set to 32000 Hz , 16 bit , mono .

Speech material was gathered in the lab setting in the context of mother–child free-play situations with 28 puppets with pre-assigned names written on them. The mother was instructed to "teach" the child the names of puppets, and the elicitation

strategies were oriented to obtain spontaneous sound productions. The children not only imitated immediately, but also produced the $C_1V(:)_1C_1V(:)_1$ -names in deferred imitation (that is, near-spontaneous speech). Caregivers were instructed to have their child produce each puppet name at least five times. Children's attempts at producing the vowels in the bi-syllabic names of the puppets (e.g., 'pi:pi:', 'ka:ka:') were judged to be either correct or incorrect in relation to the mother's target utterance by an independent naïve monolingual native speaker–listener. Reliability was examined by obtaining vowel quality judgments from a second naïve native speaker of Hungarian on 12% of the tokens. Overall reliability for the two-year-olds was 80.9%. The data set for the two-year-old Hungarian boys consisted of 25 types out of the 28 possible puppet names. The range of different puppet names per child varied between 17 and 25, and sometimes a child produced a specific puppet name only once. From the recordings of caregiver–infant dyads, these "imitated" puppet name segments of the children were selected and subjected to automated acoustic analyses. The number of utterances produced by the 8 Hungarian-speaking boys was 763.

First, children's production data were analyzed in each language separately (see Section 4.1). Next, for the comparison of the two languages, the data sets were balanced by selecting an equal number of utterances in each language (see Section 4.2).

3. Acoustic analyses

Estimating formants objectively in high-pitched sounds is notoriously difficult. The vocal tract acts as a filter for the sound source. At the same time, the source and filter functions, strictly spoken, cannot be separated since the signal is the result of both functions. Assuming the spectral envelope of the source (after correction for its slope) is being flat, the vowel quality information is completely contained in the spectral envelope of the microphone signal. High pitch causes undersampling of this spectral envelope, thus limiting the ability to detect peaks from low formants and formants close to one another. Many procedures have been

developed for formant extraction, of which the LPC inverse filtering method is widely accepted. However, as we showed in Fig. 1, wrongly chosen parameters could easily produce seriously misleading results (e.g. Palethorpe et al., 1996; Wempe, 2001; Vallabha and Tuller, 2002). High fundamentals aggravate this problem.

The current study utilizes a pitch-related band-filter analysis (Wempe and Boersma, 2003) via scripts for the PRAAT-program. The method is named “recirculation method” since multiples of only a single period of the fundamental frequency are concatenated up to a duration of 50 ms. The spectral envelope of that sound is calculated by Gaussian bandfiltering in the frequency domain with a bandwidth higher than the fundamental frequency, thus representing the spectral distribution at that point in the utterance. Advantages of the one-period recirculation method are that the analysis can select more measurement points per utterance than e.g. a 50-ms window average spectrum method, and that averaging over fast signal changes occurring within the window can be avoided easily.

3.1. Selection criteria

The automated analyses of young children’s utterances are ultimately aiming at an estimation of spectral envelope representations of the utterance. Several pilot studies were carried out examining the acoustic characteristics of infant’s utterances (e.g. F_0 and the intensity range in the Dutch data) to decide the suitable parameters for selection and measurement. Reported in previous work (Van der Stelt et al., 2003a), the distribution of fundamental frequency (F_0) was analyzed in all voiced utterances of hearing and hearing-impaired children from 5 months up to two years of age. In these young children, the F_0 -range was considerable (between 100 and 800 Hz). The mean F_0 also varied in individual children and in different recordings, but remained below 425 Hz in most cases. This value determined the bandwidth chosen for the filters used in this research, which was therefore set to 1.1×425 Hz, i.e. slightly higher than the F_0 , which appeared to be a good compromise between frequency resolution and F_0 –“rip-

ple”. For that reason, only utterance parts with an F_0 below 425 Hz were selected (see Section 3.3).

It was also necessary to avoid clipped utterances (whenever present) so that the effects of possible wrongly adjusted recording levels are minimized. Finally, possibly consonantal parts of the utterance were excluded as much as possible by omitting low amplitude levels relative to the absolute peak level. Therefore, specified intensity criteria in the script were required (see Section 3.2).

3.2. Selection procedure

The sound files used for the vowel measurements consisted of all individual utterances of the children from both language communities. The PRAAT script was told where to find the audio files. First, each vocalization was divided into 10 parts of equal duration, which permitted the distribution of measurement positions over the entire utterance, thus covering a possible articulatory change within that vocalization. The bandpass filter analysis, described below, was carried out in all 10 parts of each sound, as long as F_0 was below 425 Hz for at least 5 contiguous pitch frames. Additionally, to avoid clipping the maximum allowed amplitude level was chosen to be < -0.5 dB relative to the absolute peak level within the utterance and to avoid possibly consonantal parts of the utterance the amplitude level had to be > -10 dB relative to the absolute peak level. The script allows for displaying selected measurement points in each utterance’s oscillogram. For checking purposes, a short part around the selected point of the utterance can be made audible as well. At those points in time where all criteria are met the analysis algorithm was applied. Fig. 4 depicts the block diagram displaying the steps and loops in the automated selection procedure of vocalization parts with the required parameters. The selection procedure represents one module in the automated acoustic analysis (see Fig. 5).

3.3. Analysis method

In principle, the spectral information is embedded in one single F_0 period. In each part of an utterance that meets the selection criteria the near-

Selection Procedure (Praat script)

Fig. 4. Block diagram depicting the PRAAT script selection procedure that was used to determine the acoustic criteria for the selection of the targeted sound fragments.

est F_0 period was selected at the position of the middle frame of 5 contiguous pitch frames. This period was then recycled up to a duration of 50 ms. Recirculation of a single period has been done before in phonetic research: Pols (1977) utilized this approach in his dissertation and for research done earlier. Here, we used it as a tool for adapting the sound for our frequency-domain bandfiltering.

This constructed sound was multiplied with a Kaiser-window (which suppresses the 20 Hz ripple in the spectrum caused by the 50 ms window length). Then, pre-emphasis was applied. As a next step, a swept Gaussian bandpass filter analysis (step = 175 Hz, effective bandwidth = 1.1×425 Hz) was carried out (see Section 3.1 for the bandwidth criterion). Furthermore, a level normalization (to 0.3 Pa Root Mean Square) was applied to each recycled sound, which avoided energy variance caused by (sometimes big) differences in recording levels. If desired, a spectral intensity contour can be plotted representing a bandfilter spectrum covering a range from 0 to 7 kHz. The frequency range for our analyses was purposefully limited to the range between 0 and 7000 Hz, being sufficiently high for the children's vowel sounds, and also limiting the possible influences due to noisy recording situations. Linear rather than logarithmic intensity values of the filter measurements were used for further analysis, because this approach seemed to produce better interpretable data, as suggested by preliminary results. The 175 Hz steps result in 40 values in each spectral measurement.

Ideally, 400 values can be calculated (10 parts $\times$ 40 bins) and collected in each utterance. Due to the selection criteria for pitch and intensity (see Section 3.1 and 3.2) the average number of measurements per utterance was 5.2 for the Dutch data set and 4.8 for the Hungarian set (see Tables 5 and 6 in Section 4.2) For each language, a matrix was produced representing the intensity in each of the 40 filters for the total number of measurements. The number of rows in the matrix represented the number of measurement points for the whole data set. The 40 columns represented the bins of the bandfilter spectra covering the range from 0 to 7 kHz. In the spectra, formant-like max-

*Selection and Analysis Algorithm
(Praat script)*

Fig. 5. Block diagram depicting the PRAAT script algorithm that was utilized for the analysis of the Dutch and Hungarian data.

ima could be detected automatically, but the number of formant peaks varied greatly over the different utterances. This variability made comparison between individual envelopes difficult. In addition, the intensity information was utilized as well, while peak estimation (F_1 and F_2 selection) ignores intensity level differences. Therefore, the whole-spectrum approach was used and a subsequent data reduction via a Principal Component Analysis (PCA) was carried out. Each row in the band-filter matrix then could be regarded as a point in a 40-dimensional space. By using the first two principal components that explained the maximal amount of variance in these data points, the data were reduced and could be displayed as a pc1–pc2 plane. These results were interpreted as being similar but not necessarily identical to a (rotated) F_1 – F_2 plane (e.g. Pols, 1977).

The automated analysis method presented here appears to be a reliable approach for approximation of the spectral distribution, especially since it can successfully handle the effects of the high F_0 and avoids “wrong” spectra that sometimes occur when using the more widely used methods (see Section 1.2).

4. Results

4.1. Projecting the corner vowels in the reference planes

In an attempt to interpret the measurement results on unlabeled utterances of 2-year-old boys depicted in the pc1–pc2 planes, corner vowels from the utterances in the children’s sets were selected. This is the point where the perception/interpretation of the researchers comes into account. Two-year-old children’s repertoire is not easy to be categorized with regard to vowel quality. The first (Dutch) and the second (Hungarian) author judged the sounds without relying on any context, which made this task even more challenging. Arguably, the few corner vowels that were selected probably represent “good acoustic” examples. It was assumed, for the time being, that the chosen corner vowels in each language represented the correct acoustic vowel quality. A perception test

fell outside the scope of the current research project.

From the set of 250 unlabeled Dutch utterances, the first author selected 30 corner vowels. These utterances were analyzed and projected onto the Dutch reference plane. The 10 “good” /a:/ productions resulted in 76 measurement points, the 9 “good” /i:/s in 54, and the 11 “good” /u:/ productions in 68 measurement points. The second author selected 4 Hungarian /a:/ vowel productions, 5 /i:/ vowels, and 9 /u:/ vowels that were perceived to be produced correctly. Analysis of these vowel sounds resulted in 15, 22, and 58 measurement points, respectively.

The gray plusses in the reference planes (see Figs. 6 and 8 for example) represent the spectra from the measurement points in the unlabeled utterances. In the figures all corner vowels are represented orthographically by the a, i, and u characters. Both for Dutch and Hungarian the /a:/ and /i:/ were long vowels, for Dutch the /u:/ was a long vowel, while for Hungarian the short /u/ vowel was selected.

Fig. 6. Corner vowels projected onto the vowel space reference plane for Dutch two-year-old boys. Ellipses represent 1 s.d. distance from the mean of the measured variance for each vowel: /a:/ in red, /i:/ in green, and /u:/ in blue. Number of measurement points (gray+) is 1083 in 250 utterances. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 7. Graphs of the first two eigenvectors of a PCA of the 1083 measurement points found for the Dutch data set, explaining 54.2% of the total variance.

Fig. 8. Corner vowels projected in the vowel space reference planes for Hungarian two-year-old boys. Ellipses represent 1 s.d. distance from the mean of the measured variance for each vowel: /a:/ in red, /i:/ in green, and /u/ in blue. The number of measurement points (gray+) is 3277 in 763 utterances. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 9. Graphs of the first two eigenvectors of a PCA of the 3277 measurement points found for the Hungarian data set, explaining 58.6% of the total variance.

first 2 PCA eigenvectors are given graphically for the Dutch and Hungarian data, respectively. Tables 3 and 4 show the percentages of explained variance for the first 4 eigenvectors for the Dutch and Hungarian data.

The Hungarian reference plane has approximately 3 times more measurement points than the Dutch reference plane. This difference is due to the different numbers of utterances in the data sets.

4.2. Comparing linguistic backgrounds

As stated previously, the Dutch and Hungarian adult vowel systems are supposed to be similar with regard to several articulatory aspects. In or-

Table 3
Percentage of variance explained by each of the first 4 eigenvectors in the PCA's for all analyzable utterances of the two-year-old Dutch boys

Eigenvector 1	29.9%	29.9%
Eigenvector 2	24.3%	54.2%
Eigenvector 3	10.1%	64.3%
Eigenvector 4	9.5%	73.8%

Table 4
Percentage of variance explained by each of the first 4 eigenvectors in the PCA's for all analyzable utterances of the two-year-old Hungarian boys

Eigenvector 1	32.9%	32.9%
Eigenvector 2	25.7%	58.6%
Eigenvector 3	10.1%	68.7%
Eigenvector 4	6.5%	75.2%

der to construct a valid interpretation of the results in both languages, the creation of a common reference plane was required. To achieve this goal, a representative subset from the Hungarian data set was selected. The number of Dutch utterances in the data set that was analyzed was 229 from a total set of 250. An equal number of vocalizations was selected from the Hungarian tokens, resulting in similar sample sizes from both languages. The Hungarian set consisted of 28-target words in the form of bi-syllabic puppet names like ‘bibi’ (see Section 2.3). The two-year-olds realized 25 different target words with a total of 763 utterances. The range of produced puppet names per child varied from 17 to 25. Occasionally, a child produced the puppet’s name as many as 7 times in a form that was acceptable for the independent native native listener-judge. For each puppet name realization 9 utterances were selected randomly from the data set recorded from the 8 boys. An additional 4 utterances from the /pi:pi:/ and /bubu/ sets were utilized, since these two types were present in the data set in much larger numbers than the other types. In this manner a total of $9 \times 25 + 4 = 229$ Hungarian utterances were selected. All these Hungarian utterances met the selection criteria. The analysis of these utterances resulted in 988 measurement points.

For the Dutch data set, 1083 measurement points were identified from 229 acceptable utterances. Applying a new PCA to the 2071 (988 + 1083 combined Dutch and Hungarian) measurement points created the Dutch–Hungarian reference plane. In Tables 5 and 6 the mean numbers and standard deviations are given for the number of measurements points per child, as well as the number of utterances that was acceptable with regard to the selection criteria.

The corner vowels (see Section 4.1 for the selection method) produced by both groups of children were mapped onto the common reference plane (see Fig. 10 for the Dutch vowels, and Fig. 11 for Hungarian vowels). Table 7 shows the percentages of explained variance for the first 4 eigenvectors of the PCA of the balanced Dutch–Hungarian data sets. Fig. 12 depicts the graphs of the first two eigenvectors.

Globally, the extent of the vowel utterances produced in both corpora is quite similar in the common space. This is in good agreement with the similarities of the two considered vowel systems. The differences in selected “corner” utterances are not easy to interpret. They could actually be due either to differences in languages, in corpora acquisition, or even in the selection done by the two authors.

Table 5

Number of utterances and measurement points per child, as well as the average number and s.d. in the recordings of the 5 two-year-old Dutch boys

Dutch	Boy 1	Boy 2	Boy 3	Boy 4	Boy 5	Total
Valid utterances from 50	36	49	49	47	27	208
Total of selected positions	119	322	267	248	127	1083
Average number per utterance	3.3	6.6	5.4	5.3	4.7	5.2
s.d. number per utterance	1.7	2.1	2.0	2.4	1.6	

Table 6

Number of utterances and measurement points per child, as well as the average number and s.d. in the recordings of the 8 two-year-old Hungarian boys

Hungarian	Boy 1	Boy 2	Boy 3	Boy 4	Boy 5	Boy 6	Boy 7	Boy 8	Total
Valid utterances	46	17	24	20	21	39	23	18	208
Total of selected positions	269	70	105	88	78	161	120	97	988
Average number per utterance	5.8	4.1	4.4	4.4	3.7	4.1	5.2	5.4	4.75
s.d. number per utterance	2.2	1.6	1.6	1.7	1.4	1.7	2.2	1.1

Fig. 10. Common Dutch–Hungarian reference plane created by means of the first two eigenvectors. The gray plusses (+) representing the spectra of measurement points in $2 \times 229 = 458$ utterances: all together 2071 points. The Dutch /a:/-vowels that were perceived to be correctly produced are given red, the /i:/-vowels in green, and the /u:/-vowels in blue. Ellipses represent 1 s.d. distance from the mean of the measured variance for each vowel. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 11. Common Dutch–Hungarian reference plane created by means of the first two eigenvectors. The gray plusses (+) representing the spectra of measurement points in $2 \times 229 = 458$ utterances: all together 2071 points. The Hungarian /a:/-vowels that were perceived to be correctly produced are given red, the /i:/-vowels in green, and the /u:/-vowels in blue. Ellipses represent 1 s.d. distance from the mean of the measured variance for each vowel. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

Table 7

Percentage of variance explained by each of the first 4 eigenvectors in the PCA for the 458 analyzable utterances of the two-year-old Dutch and Hungarian boys together

Eigenvector 1	28.0%	28.0%
Eigenvector 2	24.1%	52.1%
Eigenvector 3	13.6%	65.7%
Eigenvector 4	10.0%	75.7%

4.3. Various effects of data selection on the reference plane

4.3.1. Contribution of the number of utterances and measurement points

Results displayed in Tables 5 and 6 suggest that the original number of utterances and the number of measurement points for each child and each language varied largely. Since the number of measurement points might have an effect on the shape of the reference plane, the positions of all measure-

ments from the second Dutch subject and the positions of a randomly taken half of those measurements were compared. Fig. 13 displays the common reference plane with both one-sigma ellipses, for the total number of 322 measurement points (red ellipse) as well as for half the number of measurement points (161, the green ellipse). It can be concluded that even in the case where the number of measurement points varied a factor two, there is only a small effect on the spread of measurement points in the display.

4.3.2. The effect of background noise on the measurements

The effect of the noise levels on the measurements needed to be checked. The good Dutch /a:/ productions were selected for this check. Pink noise (with a constant level of energy per octave) was added to the recirculated sound via the PRAAT script. The noise level was -40 dB with

Fig. 12. Graphs of the first two eigenvectors of a PCA of the 2071 measurement points found for the balanced Dutch–Hungarian data set of 458 utterances, explaining 52.1% of the total variance.

Fig. 13. Contribution of the number of measurement points to the spreading over the reference plane. The green forward slashes give all 322 measurement points for the second Dutch child. The red back slashes indicate the randomly selected half of the measurement points. The green ellipse represents 1 s.d. distance from the mean of the measured variance for all these 322 points (see Table 5). The red ellipse represents 1 s.d. distance for only 161 measurement points. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 14. The effect of pink noise added to the recirculated sounds. The data set consisted of the Dutch good /a:/ productions. The green forward slashes represent the measurements without noise, the red backward slashes are the measurements when -40 dB pink noise is added. The red ellipse completely covers the green ellipse, both representing 1 s.d. distance from the mean of the measured variance for noisy and non-noisy sounds, respectively. (For interpretation of the references in colour in this figure legend, the reader is referred to the web version of this article.)

with noise. Projecting both types of slashes results in crosses, demonstrating that a signal-to-noise ratio of -40 dB does not change the measurements. Both ellipses for 1 s.d. distance from the mean of the measured variance in the noisy and non-noisy /a:/ measurements (green and red, respectively) completely coincide. Therefore, only the last drawn ellipse drawn (red) can be detected. Thus, it is argued that the signal-to-noise ratios of the recordings (see Tables 1 and 2) allow for reliable measurements.

5. Discussion

At first glance, it may seem strange to talk about children's vowel spaces while utilizing measurements of their unlabeled vowel sounds. Usually, the researcher puts a lot of effort into identifying the right types of utterances, and then

reference to the level of the recirculated sound. The spectral points are given for the /a:/ measurements without and with the pink noise in Fig. 14. The forward slashes represent the points without noise; the backward slashes represent the points

selecting and categorizing the sounds in order to be able to answer the research questions. Hand-edited acoustic measurement techniques are typically subjective and very time consuming.

Our approach to identifying patterns in two-year-old children's vowel spaces started the exploration process from a different angle. The large amount of data to be examined necessitated the use of automated analysis. The Dutch audio material of 5 hearing and 5 deaf babies was collected monthly between birth and the 18th month of life, with an additional recording at 24 months. The total set of Hungarian material that we intend to examine by this method at a later moment in time is even larger: 8 boys and 8 girls in each age group, at the ages of 2;0, 2;6, 3;0, 3;6 and 4;0 years old. In the current study, we only examined data from the Dutch and Hungarian two-year-old normally hearing boys.

It is clear that for language comparison, the methods for collecting utterances must be similar as much as possible. The different spreading found per language, very likely stems at least partly from the different methods eliciting the utterances. By using existing corpora, the problem that acoustic differences are difficult to compare and interpret has to be faced. Reassuring is the finding that manipulation of the number of utterances and the number of measurement points (if sufficiently large, see Tables 8 and 9) does not severely affect the spreading over the reference plane.

In spite of the care taken when obtaining audio recordings the quality of recordings often end up being poor. Recording children's speech is typically an even greater challenge than recording speech from adults: the unpredictable volume of children's voice and the fact that they move around (away from the microphone) present the researcher with additional problems. The criteria for the selection and analysis method presented here appear to be robust, even with regard to poor signal-to-noise ratios. Especially for children's data this is a desirable result.

Surely, the development of the analysis method is not finished yet. However, at this point the authors are pleased that the unlabeled reference plane for Dutch- and Hungarian-speaking children's utterances appears to be useful as far as

the spectral envelopes of labeled utterances (/a:/, /i:/, and /u:/ vowels) show clustering when projected on the reference plane. Large amounts of unlabeled data can now be collected and then examined in the context of measurements taken from a few precisely labeled tokens.

Relating the acoustic reference plane to the well-known F_1 - F_2 (articulatory) planes of various languages is possible via vowel synthesis. With an articulatory model, such as The Variable Linear Articulatory Model (VLAM) (Boë and Maeda, 1997; Ménard et al., 2002), it is possible to construct baby sounds with known formants, as well as known articulatory characteristics. Analyzing these artificial sounds by means of the recirculation method facilitates the interpretation of the reference plane. Since the VLAM permits age manipulation as well, the effect of growing articulators on vowel formant measures can also be studied. Even with only F_1 - F_2 values, the pertaining sounds may be reconstructed to some extent, and thus comparison of that F_1 - F_2 plane with the reference plane found by bandfiltering the reconstructed sounds becomes possible.

Similarly to the current research project, where the second author observed that the Hungarian phonemically short /u/-vowel is produced by relatively strong lip rounding in some (but not all) two-year-olds, it is hoped that further correct classification of vowel production in young children will demonstrate the validity of interpretation of those results projected onto the reference plane. As a next step, the exploration of less well-articulated vowels needs to follow.

6. Conclusion

Clearly, the automated bandfilter analysis approach holds great promise for opening up a new venue for studying acoustic aspects in the vowel productions of young children. Via this approach, important similarities and differences in vowel acquisition patterns in various languages can be uncovered. An imperative next step is unraveling the relations between the acoustic bandfilter approach and the more articulatory oriented LPC analysis. The results may aid us in identifying ele-

ments of a unified theory of vowel acquisition. Since a theory of vowel acquisition does not exist today, any progress we can make by utilizing the current method to map out stages and processes of vowel acquisition in children will add a new building block towards creating such a theory.

7. Uncited references

(Van den Dikkenberg-Pot and Koopmans-van Beinum, 1997).

Acknowledgement

Funding for this project for the second author is gratefully acknowledged from the College of Health Sciences at the University of Wyoming. Additional funding was also provided in the form of an “International Travel Grant” from the International Programs Office at the University of Wyoming; without this much appreciated award, the cooperation between the University of Amsterdam and the University of Wyoming would not have been possible. The second author thanks whole-heartedly two graduate student assistants: Jody Olstad and Janine Hilsher, at the Division of Communication Disorders of the University of Wyoming who devoted much of their time to the selection and saving of tokens.

The major inspiration behind this study is the early work undertaken by Prof. Dr. I.r. Louis C.W. Pols (e.g., 1977) to explore vowel production characteristics.

Without his intellectual support, this article would not have been written. Louis, in fact, should have been one of our co-authors. During the earliest phases of this project, we consulted him about several challenging questions that arose. Louis encouraged us to develop this study into a series of experiments, not knowing that the article was meant to appear in his farewell “Festschrift”. Such things happen when preparing a surprise. We have been constantly in touch with the editors about our strategy. Presenting the reviewers’ comments to Louis urged us to use our poker faces again.

So a whole-hearted thank you, Louis, for all your critical interest, encouragement and help. We promise to continue exploring this field further in your honour.

References

- Boë, L.-J., Maeda, S., 1997. Modélisation de la croissance du conduit vocal. Espace vocalique des nouveau-nés et des adultes. Conséquences pour l’ontogenèse et la phylogenèse, Journées d’Études Linguistiques: “La Voyelle dans Tous ses États”, Nantes, pp. 98–105.
- Boersma, P., Weenink, D., 1996. PRAAT: a system for doing phonetics by computer, version 3.4., Institute of Phonetic Sciences of the University of Amsterdam, Report number 132.
- Clement, C.J., 2004. Development of vocalizations in deaf and normally hearing infants, PhD dissertation, University of Amsterdam.
- Davis, B.L., MacNeilage, P.F., 1990. Acquisition of correct vowel production: a quantitative case study. *J. Speech Hear. Res.* 33, 16–27.
- De Boysson-Bardies, B., Halle, P., Sagart, L., Durand, C., 1989. A crosslinguistic investigation of vowel formants in babbling. *J. Child Lang.* 16, 1–17.
- Kent, R.D., 1993. Infants and speech: seeking patterns. *J. Phonetics* 21, 117–123.
- Koopmans-van Beinum, F.J., Doppen, L., 2003. Development in utterance structures of deaf and hearing infants. In: Solé, M.J., Rescasens, D., Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona* 1, pp. 1033–1036.
- Koopmans-van Beinum, F.J., Van der Stelt, J.M., 1986. Early stages in the development of speech movements. In: Lindblom, B., Zetterström, R. (Eds.), *Precursors of Early Speech*. Stockton Press, New York, pp. 37–50.
- Kuhl, P.K., 1993. Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *J. Phonetics* 21, 125–139.
- Kuhl, P.K., Meltzoff, A.N., 1982. The bimodal perception of speech in infancy. *Science* 218, 1138–1141.
- Kuhl, P.K., Meltzoff, A.N., 1996. Infant vocalizations in response to speech: vocal imitation and developmental change. *J. Ac. Soc. Am.* 100, 2425–2438.
- Lee, S., Potamianos, A., Narayanan, S., 1997. Analysis of children’s speech: duration, pitch and formants. In: *Proceedings Eurospeech ’97*, pp. 473–476.
- MacNeilage, P.F., 1998. The frame/content theory of evolution of speech production. *Behav. Brain Sci.* 21, 499–546.
- Ménard, L., Schwartz, J.L., Boë, L.-J., Kandel, S., Vallée, N., 2002. Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *J. Ac. Soc. Am.* 111, 1895–1992.

- 903 Palethorpe, S., Wales, R., Clark, J.E., Senserrick, T., 1996. 936
 904 Vowel classification in children. *J. Ac. Soc. Am.* 100, 3843– 937
 905 3851. 938
- 906 Plomp, R., Pols, L.C.W., Van de Geer, J.P., 1967. Dimensional 939
 907 analysis of vowel spectra. *J. Ac. Soc. Am.* 41, 707–712. 940
 908 Pols, L.C.W., Van der Kamp, L.J.Th., Plomp, R., 1969. 941
 909 Perceptual and physical space of vowel sounds. *J. Ac. Soc.* 942
 910 *Am.* 46, 458–467. 943
- 911 Pols, L.C.W., Tromp, H.R.C., Plomp, R., 1973. Frequency 944
 912 analysis of Dutch vowels from 50 male speakers. *J. Ac. Soc.* 945
 913 *Am.* 53, 1093–1101. 946
- 914 Pols, L.C.W., 1977. Spectral analysis and identification of 947
 915 Dutch vowels in monosyllabic words. Ph.D. dissertation, 948
 916 Free University of Amsterdam. 949
- 917 Robb, M.P., Chen, Y., Gilbert, H.R., 1997. Developmental 950
 918 aspects of formant frequency and bandwidth in infants and 951
 919 toddlers. *Folia Phon. Log.* 49, 88–95. 952
- 920 Serkhane, J., Schwartz, J.L., Boë, L.J., Davis, B., Matyear, C., 953
 921 2002. Motor specifications of a baby robot via the analysis 954
 922 of infants' vocalizations. In: Proceedings of the Interna- 955
 923 tional Conference on Spoken Language Processing, Denver, 956
 924 pp. 45–48. 957
- 925 Stoel-Gammon, C., 1983. Constraints on consonant-vowel 958
 926 sequences in early words. *J. Child Lang.* 10, 455–457. 959
- 927 Van den Dikkenberg-Pot, I., Koopmans-van Beinum, F.J., 960
 928 1997. Number and duration of spoken utterances of deaf 961
 929 and normally hearing children between 12 and 18 months. 962
 930 In: Proceedings of the Institute of Phonetic Sciences of the 963
 931 University of Amsterdam, Vol. 21, pp. 91–100. 964
- 932 Van der Stelt, J.M., Wempe, A.G., Pols, L.C.W., 2003a. 965
 933 Progression in infants' vowel space: an analysis of deaf and 966
 934 hearing infants' sounds. In: Solé, M.J., Rescasens, D., 967
 935 Romero, J. (Eds.), Proceedings of the 15th International 968
 Congress of Phonetic Sciences, Barcelona, Vol. 3, pp. 2225– 936
 2228. 937
- Van der Stelt, J.M., Wempe, A.G., Pols, L.C.W., 2003b. 938
 Progression in vowel production: comparing deaf and 939
 hearing children. In: Proceedings of the Institute of Phonetic 940
 Sciences of the University of Amsterdam, Vol. 25, pp. 197– 941
 206. 942
- Vallabha, G.K., Tuller, B., 2002. Systematic errors in the 943
 formant analysis of steady-state vowels. *Speech Comm.* 38, 944
 141–160. 945
- Wempe, A.G., 2001. F0-related formant measurements. In: 946
 Proceedings of the Institute of Phonetic Sciences of the 947
 University of Amsterdam, Vol. 24, pp. 167–187. 948
- Wempe, A.G., Boersma, P., 2003. The interactive design of an 949
 F0-related spectral analyser. In: Solé, M.J., Rescasens, D., 950
 Romero, J. (Eds.), Proceedings of the 15th International 951
 Congress of Phonetic Sciences, Barcelona, Vol. 1, pp. 343– 952
 346. 953
- Zajdó, K., 2002a. Vowel acquisition in Hungarian: a first look 954
 at the data. In: Larson, J., Paster, M. (Eds.), Proceedings of 955
 the Twenty-Eighth Annual Meeting of the Berkeley Lin- 956
 guistics Society. Berkeley Linguistics Society, Berkeley, CA, 957
 pp. 363–374. 958
- Zajdó, K., 2002b. The acquisition of vowels by Hungarian- 959
 speaking children aged two to four years: a cross-sectional 960
 study. Unpublished doctoral dissertation, University of 961
 Washington, Seattle, USA. 962
- Zajdó, K., Stoel-Gammon, C., 2003. The acquisition of vowels 963
 in Hungarian: developmental data. In: Solé, M.J., Resca- 964
 sens, D., Romero, J. (Eds.), Proceedings of the 15th 965
 International Congress of Phonetic Sciences, Barcelona 3, 966
 pp. 2229–2232. 967
 968