ABSTRACT

In this paper, the phonetic aspects of babbling and early words are investigated through auditory and acoustic analysis. Two children were recorded at 10, 12, 14, and 16 months of age during play situations in the presence of their mothers. Every occurrence of babbling episodes and early words suitable for acoustic analysis were classified for consonantal and vocalic categories according to IPA symbols. Acoustic analysis was then performed on the vowels to describe their F1 and F2 values. The acoustic values of the vowels were compared longitudinally with the available data for a representative sample of 7-year-old Italian children and adults (males and females).

1. INTRODUCTION

The great individual variability of the speech development during babbling and the first words stage has been highlighted by a number of authors (cf. [1]).

The key issue is represented here by the complex time-course of the progressive disengagement from the biomechanical constraints grounded in the canonical babbling towards the achievement of the phonetic mastery for the first phonological oppositions. This process of development involves a large number of variables of different nature, both internal to the child (biological time-schedules) as well as external (the language spoken in the environment), strictly interacting together.

The stages of this evolution may be depicted essentially in two ways: either through the use of frequency statistics based on auditory classifications of the tokens along the phones’ categories, or through a measurement of the physical properties of the signals, such as formant frequencies (Hz), dependent on the anatomic shape and dimension of the vocal apparatus and the mastery in its use, and durations (s). Although temporal measurements were also taken in the signals studied here, we concentrated only on frequency statistics of phones and CV associations, and on F1 and F2 measurements.

2. METHOD

2.1. Subjects and Materials. The two children attending the study were Davide (a male) and Caterina (a female). They were born full-term, and their course of development were reported to be normal and free of any specific problems relating to their speech and hearing apparatus. Their parents are middle-class, and the language spoken is Italian. The children were recorded every two weeks from the 8th month onwards, during play situations and in the presence of their mothers. The recordings, lasting 45 minutes each, were made at their respective home in a noise-free environment. Davide was recorded with an AIWA tape recorder, mod. TP-850 and a Sony microphone, mod. ECM-T7, whereas Caterina was recorded with a Sony DAT recorder, mod. TCD-D8 and a Sony microphone, mod. ECM M5 907. The analysis of only 4 recordings are presented here, referring to the 10th month (9;09 and 9;28, months and days, for Davide; 9;05 for Caterina), 12th month (11;21 for Davide; 11;12 for Caterina), 14th month (13;07 for Davide; 13;14 for Caterina) and 16th month (15;16 for Davide; 15;14 for Caterina). According to the MacArthur Instrument [2], Davide were producing 4 words at 9;09, and 31 words at 13;01, whereas Caterina were producing 6 words at 11;6 and 13 at 12;16.

2.2. Preliminary transcription and tokens’ selection. A first phonetic transcription, performed by the first author using the symbols and diacritics provided in [3] besides those of the IPA (1996), was the basis to select babbling and word occurrences, that were acquired at 25.6 kHz sampling frequency with the CSL 4300 package and examined for compliance with the acoustic criteria proposed by Oller [4]. In this manner, it was possible to exclude any noised production and discriminate babbling and any words from vegetative and reflexive sounds as well as from sounds belonging to the vocal play stage. Isolated CV syllables were accepted, but only vowels produced with loosely defined “modal” voice (i.e. neither laryngalized, strongly nasalized, desonorized, nor in falsetto) and preceded by “true” consonants (no glottal consonants) were analyzed. The occurrences characterized by F0 shift, harmonic doubling, biphonation [5] were also tentatively individuated and excluded. This revision resulted in the elimination of 422/1172 potential babbling utterances. Two successive syllables were considered to refer to different utterances if separated by more than 500 ms. In the presence of vocalic modulations, the number and type of occurring vowels was specified, keeping in mind the possibility that a vowel may have very long transitions.

2.3. Final transcription. A new phonetic transcription was performed by the first author on the selected occurrences, exploiting the facilities of the CSL 4300 package and an interactive phonetic symbol guide [6]. Further, the whole process relative to the final phonetic classification was aided by the comparison with an interactive archive of the stored vowels which had received the same classification by 4 independent transcribers (at least 2 tokens for each vowel). The validity of this transcription was checked with two different procedures. For Davide, the first author and an expert transcriber worked jointly on the 25 % of the whole corpus, and discussed every disagreement, attempting to achieve a common classification. The final agreement score for the IPA symbols was quite high (98,3%). As for Caterina, the agreement’s percentage of 3 out 4 independent judges on a sample of 168 vowels, randomly chosen across all the recordings, was 31.5% (40.4% not considering the tract “rounded/unrounded”). The percentage score for the consonants was calculated on the transcriptions, performed by 2 independent experts, of randomly selected samples of 10
consecutive utterances for each recording. By applying the formula in [7], the resulting score was 59.1%.

2.3. Acoustic analysis. The formant analysis on vowels was performed on the most stable part (i.e. where F1 and F2 were possibly flat, clearly discernible and separated from each other). For intra-subject agreement, the first 100 LPC analysis (14 poles, sometimes 16 poles), averaged on these portions, was compared with a manual interpolation of the harmonic envelope (50 Hz bandwidth) on the same occurrences. As the results of a t-test did not reveal any significant difference between the two series of measurements, preference was given to the automatic LPC measurements, except for the evidently wrong cases. The inter-subject agreement was checked on 10% of the whole corpus with a t-test, measures, except for the evidently wrong cases. The inter-subject measurements, preference was given to the automatic LPC measurements, except for the evidently wrong cases.

3. RESULTS

3.1. Some quantitative characteristics of the corpus. As a general note, the results pertaining to Davide, though already published in [8], will be presented and discussed here where needed. Table 1 shows statistics describing some quantitative characteristics of the analyzed corpus.

<table>
<thead>
<tr>
<th>months</th>
<th>No. babbling and words utterances</th>
<th>No. of analyzed syll. in babbling (in words)</th>
<th>No. of analyzed syllables in mono-/plurisyllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>10th</td>
<td>Davide 137328 Davide 140</td>
<td>Davide 137328 Davide 140</td>
<td>Davide 137328 Davide 140</td>
</tr>
<tr>
<td>12th</td>
<td>283480 241490</td>
<td>241490</td>
<td>241490</td>
</tr>
<tr>
<td>14th</td>
<td>81182  155411</td>
<td>155411</td>
<td>155411</td>
</tr>
<tr>
<td>16th</td>
<td>122163 190342</td>
<td>190342</td>
<td>190342</td>
</tr>
</tbody>
</table>

Table 1. Counts of babbling utterances and words over total vocalizations, counts of the analyzed syllables (those belonging to words are enclosed between parenthesis) and counts of the analyzed syllables belonging to either monosyllables or plurisyllables.

The total number of syllables analyzed are 359 for Davide and 758 for Caterina. The babbling utterances and the first words increase their proportion (along time) over the total number of vocalizations (col. 1), and the syllables in words increase their proportion among all the analyzed syllables (col. 2).

3.2. Results based on auditory classifications. After classifying the types of syllables produced by the 2 subjects in terms of CV structure, their relative frequency over the total number of syllables was compared to the relative frequency of the syllable types in a selection (440 syllable tokens), performed by the first author, of the target words attested by parents as present in the vocabulary of at least 3 out of 315 children (appendix B of [2]). The two most frequent types within the target lexicon are CV (61.1%) and CVC (20.2%). The CV type also constitutes the great majority of the productions for the 2 children, ranging from 80% to 90% for Davide, and always above 90% for Caterina. However, a direct comparison between the children’s and adults’ percentages is made difficult by the fact that most final consonants of the CVC types of the adult lexical selection consists of a geminated consonant, and consonantal lengthening was never transcribed in the children’s syllables. Table 2 lists the percentage of occurrence of the most frequent vowels (above the 5% level) across the 4 stages for the 2 children, together with 2 statistics referring to standard Italian [9, 10] and a statistics referring to the lexical selection from [2].

This table clearly illustrates the common preference accorded from the beginning to frontal vowels, and the scarce presence of back vowels. However, the 2 children differentiated for their production of low vowels, as their prevalence for Davide increased and more similar to each other and to those of Italian.

![Figure 1. Frequency occurrence percentages of vowels in the productions of Davide (D) and Caterina (C) and three main sources for Italian [9, 10, 2].](image)
3.3. Acoustic description. In the Figs. 2 and 3 the vowels produced by Davide and Caterina and classified as occurrences of standard Italian are plotted in the acoustic space (Hz) defined by the F1 and F2 coordinates (any potential allophones were excluded).

Comparing these figures, a similarity in distribution is evident only in the 14th month, when both children converge towards the adult model, distinguishing well the area of existence of all the vowels, although the ways in which they reach this final point are different. Davide starts from a virtual sticking to frontal articulations, with values concentrated in the high and right quadrant of the figure, and gradually elongates first the F2 axis and then the F1 axis. The starting distribution of Caterina is more central and balanced, and in the following stages spreads harmoniously along both axes.

In figure 4, a comparison is proposed between the mean values of the cardinal vowels /a, i, u/ produced by Davide and Caterina in 2 main stages (the 10th month with the 12th month, and the 14th month with the 16th month), and the centroids of existence for the vowels of the subjects of [11]. The adult subjects are a representative sample of Italian male and female speakers of northern Italian, and the children attend the first three classes of primary school.

Although there is no systematic lowering of the formant frequencies along with increasing of the ages, the vowels produced by Davide and Caterina in the final stage are grossly in line with the centroids. These data are best interpreted by assuming a similarity in the shape of the vocal tract and a linear factor active in the proportional growth of the vocal tract for the ages considered here.

3.4. Consonant-vowel associations. Fig. 5 illustrates a comparison between frequencies of consonants pronounced by Davide and Caterina and the frequencies attested for Italian, based on the classification of the active articulators, resulting in three classes: labials (bilabials and labiodentals), apicals (dentals, alveolars, postalveolars, retroflexes), and dorsals (palatals, velars, uvulars, pharyngeals). Only the initials of the CV and CVC syllables were considered for the productions of the two children and the words listed in the selection from [2]. Even with regard to consonants, the final stages of the 2 children are similar to each other and to the adult target, although the most dramatic changes along the way are achieved this time by
Data referring to vowel and consonant occurrence frequency, though useful, cannot provide information regarding co-occurrence frequencies, and thus preferred syllable types.

The intra-syllabic hypothesis of MacNeilage and Davis’s “Frame, then content” theory [12] makes precise predictions regarding the co-occurrences of vowels and consonants within the syllable: front vowels tend to occur after labial consonants; central vowels tend to occur after labial consonants; back vowels tend to occur after dorsal consonants. Table 3 describes the relationship between the observed and expected frequencies, by reporting their ratio for each cell (the expected value is 1.0). The predicted associations are highlighted in bold character. Data are divided up into two groups (the first 2 and the last 2 stages).

Table 3. Ratio of the observed-to-expected co-occurrence frequencies for labial (L), apical (A) and dorsal (D) consonants with the front (F), central (C) and back (B) vowels. The subjects are Davide (D, above) and Caterina (C, below).

<table>
<thead>
<tr>
<th></th>
<th>Davide</th>
<th></th>
<th></th>
<th></th>
<th>Davide</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10+12</td>
<td>14+16</td>
<td></td>
<td></td>
<td>10+12</td>
<td>14+16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>0.609</td>
<td>1.486</td>
<td>0.522</td>
<td>F</td>
<td>0.262</td>
<td>1.155</td>
<td>2.161</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1.270</td>
<td>0.905</td>
<td>0.680</td>
<td>C</td>
<td>0.521</td>
<td>1.376</td>
<td>1.009</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1.067</td>
<td>0.683</td>
<td>1.714</td>
<td>B</td>
<td>0.880</td>
<td>0.715</td>
<td>2.028</td>
<td></td>
</tr>
</tbody>
</table>

In the final months evidence a non-random distribution different from the predicted one, with an exception for the apical consonants-front vowels association (overall, the strongest association). In order to ascertain the patterns of association characterizing the target lexicon, and thus its influence, a new χ² analysis was carried out on the consonants and vowels of the lexical selection from [2]. The only significant result regarded the Dorsal consonant-vowel association, where dorsals associated were more with central vowels (1.438), and less with back vowels (0.358), than expected.

4. CONCLUSIONS

This description of some characteristics of the consonants’ and vowels’ development in 2 children along the stages of late babbling and early words production aims to demonstrates the utility of an interaction between the traditional phonetic method based on auditory classifications and the method based on acoustic analysis. The results were useful for tracing individual profiles of speech development, that in these two children was discovered to be more different at the starting point than at the final point, and revealed the increasing influence of the target language. Further, these results may offer some elements and suggestions to the researchers interested in cross-linguistic comparisons. Future research will attempt to disentangle the relative contributions from the part of babbling and words.

REFERENCES