Lund University, Dept. of Linguistics
Working Papers 49 (2001), 98–101

Re-assessing the perceptual consequences of CV-transition speed[1]

Francisco Lacerda
Dept. of Linguistics, Stockholm University
frasse@ling.su.se

Abstract

This status report of the project "Barns talperception" (Children's speech perception) presents pilot experiments carried out to investigate how transition speed may affect the perception of natural and synthetic CV-syllables. The results indicate that whereas labeling of synthetic CV-syllables tends to be affected by transition speed, their natural counterparts resist manipulations of this parameter. The project attempts to investigate how different phonetic aspects may account for the reported success of language training methods based on exposure to temporally and spectrally modified speech.

1 Introduction

In two lively debated articles (Tallal et al., 1996; Merzenich et al. 1996), Tallal and her colleagues reported spectacular improvements in the language performance of Specific Language Impaired (SLI) children after an intensive language training program using temporally modified and spectrally enhanced speech. The interpretation offered by the authors was based on the notion that SLI children suffer from a neurological deficit that impairs processing of rapid speech events, such as the formant transitions typically occurring at the stop-consonant to vowel transitions in CV-syllables. Thus, to provide the SLI children with phonetic cues robust enough to survive their processing deficit, the speech signal was pre-processed by temporal and spectral transformations that were intended to compensate for the SLI children's processing deficit. The training program was reported to be very successful but there are still some unanswered methodological and theoretical issues. It is possible, for instance, that the program's success might be due to the extensive exposure to altered speech rather than to the merits of the transformation itself (Lacerda & Lindblom 1998). To examine this question, the present project aims at addressing the potential phonetic impact of each of the manipulations used by Tallal and her colleagues. And, among all the various aspects that are being addressed, we will report here the pilot experiments we used to study the significance of the temporal aspects *per se*.

2 Method

2.1 Subjects

The subjects participating in the perception tests reported here are divided in two main groups: A group of 24 second-grade children form a Swedish public school in the Stockholm area and a group of 46 students of Linguistics (first term). All the subjects were proficient speakers of Swedish and all but one reported normal hearing. The subject

[1]Research supported by Humanistiska och Samhällsvetenskapliga Forskningsrådet (HSFR)

reporting hearing deficit had nevertheless functional hearing. Of the 24 second-graders, 21 were native speakers of Swedish. The remaining three children were proficient in Swedish but had Finish, English and Persian as their first languages.

2.2 Stimuli

A set of natural utterances, [ba], [da], [ga], [wa], [ab], [ad], [ag] and [aw], produced by an adult male speaker of Swedish and synthetic [ba], [da], [ab] and [ad] syllables were used as stimuli. The transition durations of the natural stimuli were manually edited to obtain an additional series with doubled transition durations for [ba], [da], [ab] and [ad] and a series of halved transition durations for [wa] and [aw]. The [ga] and [ag] stimuli were used as foils and were not manipulated. The synthetic stimuli consisted of four series of [ba], [da], [ab] and [ad], created with transition durations of 20 ms, 40 ms, 60 ms and 80 ms.

The synthetic stimuli were 4-formant utterances generated by a parallel speech synthesizer. The formant transitions were exponential between the static vowel and the stop-consonant loci with durations defined as the time interval to reach 90% of the total transition excursion. For each stimulus, all the formants had identical transition durations. The voice source was created by a pulse train in with F_0 decaying linearly from 135 Hz to 120 Hz throughout the stimulus. The vowel's formants were set at 592 Hz, 1070 Hz, 2400 Hz and 3300 Hz. The F-patterns of the loci were defined as F_1=200 Hz, F_2=700 Hz for [b] and 1600 [d], while F_3 and F_4 were identical to the vowel's F_3 and F_4.

2.3 Procedure

Both identification and discrimination tests were conducted with these stimuli. In a first series of experiments, the school children were requested to label all the CV-utterances as [ba], [da], [ga] or [wa]. In another experiment the complete set of natural and synthetic stimuli, i.e. both CV and VC utterances, was presented for open class labeling by a group of students from the first term of General Linguistics. This group of subjects also participated in a discrimination test in which the stimuli were presented in pairs under non-optimal listening conditions, created by background of pink noise with S/N of 0 dB and 6 dB.

3 Results

3.1 Labeling by school children

To keep the test duration as short as possible, not all the stimuli were presented the same amount of times (see Table 1).

Table 1. Type and number of stimuli used in the labeling test with school children

Type	Syllable	20	40	60	80	dubble	half	unman
natural	ba					4		2
	da					4		2
	ga							2
	wa						4	2
synth	ba	3	3	3	3			
	da	3	3	3	3			

The full set of responses provided by the children is far too complex to be presented here in detail. Figure 1 shows part of the results, viewed as points in a normalized 3-dimensional space ([b]x[w]x[d]). The data indicate that the children were virtually insensitive to doubling or halving of the formant transitions of the natural stimuli. For natural [b] and [w] the data clusters at ceiling levels in spite of doubling or halving the transition speeds.

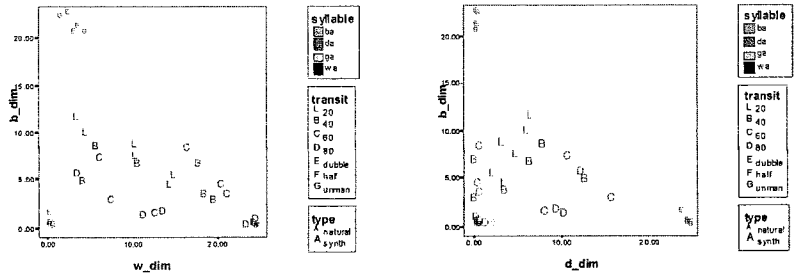

Figure 1. Children's labeling responses expressed as "b-", "d-" and "w-dimensions".

In contrast, they were sensitive to the manipulations of the synthetic stimuli. The responses to the [ba]-stimuli, for instance, land on the –45° diagonal connecting the upper left to the lower right corner of the [b]x[w] response domain (left panel), supporting the usually observed trade-off between synthetic [b] and [w] as a function of transition speed.

3.2 Free labeling by adults

The free labeling test carried out with the adult group further supports the notion that the natural CV-stimuli are virtually insensitive to manipulations of transition speeds. The four panels on the left side of Figure 2 show the label distributions generated for each of the natural CV-stimuli categories. The x-axis shows the labels chosen by the listeners. The parameter "manipl" indicates how the stimuli were manipulated ("du" stands for doubling of the transition duration, "ha", for halving and "om" for non-manipulated stimuli). Once again, the stability of the responses to these natural stimuli is rather high for all but the [ba] utterances, which tended to shift toward [ga] for the doubled transition duration. In contrast with the natural stimuli, labeling responses to the synthetic stimuli, displayed in the two panels on the right-hand side of the figure, show a substantial variation of the labels as a function of the transition speeds of the synthetic CV-stimuli. Particularly interesting is the fact that, with synthetic stimuli, [ba] responses turn into [wa] responses as the transition duration is increased, in agreement with the observations often reported in the literature, whereas the responses to natural stimuli are nearly unaffected by doubling the transition duration.

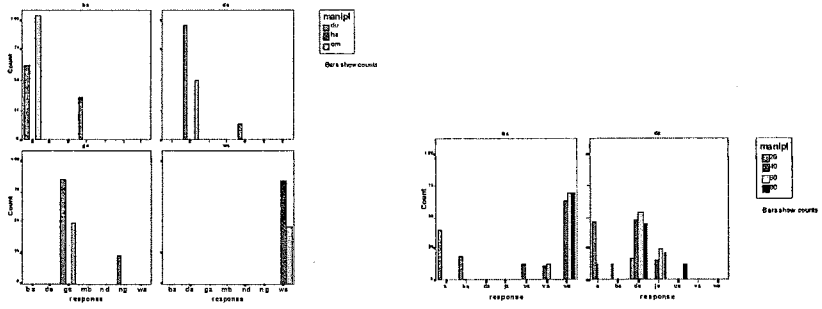

Figure 2. Adult open-class labeling results.

3.3 Discrimination of CV-syllables in noise

Figure 3 shows a summary of the percentage discrimination scores for CV-syllable pairs presented under non-optimal listening conditions. The data were pooled by the transition speeds of the pairs presented for discrimination. The panel to the left shows the results for a S/N=0 dB and the one to the right displays the results for S/N=6 dB. Along the x-axis is indicated whether the pairs judged contained the same or different intended phoneme categories. There is a reassuringly very significant difference between the "different"-judgments of the pairs containing different *vs.* same syllables. Not surprisingly, the overall performance is a little better for the higher S/N ratio. Discrimination seems to be marginally better for the "slow-transition pairs", than for the "fast-transition" ones and worse for the mixed pairs.

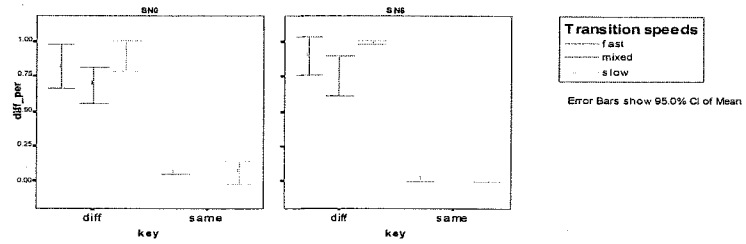

Figure 3. Adult discrimination in background noise.

4 Conclusions

The phonetic quality of synthetic CV-stimuli seems to be rather sensitive to transition speed whereas the phonetic quality of natural stimuli is nearly unaffected by similar manipulations. Discrimination results from synthetic stimuli cannot therefore be directly extrapolated to their natural models. The improvement in discrimination performance observed in SLI children tested with manipulated synthetic speech may have been due to apparent differences in phonetic quality rather then to a direct compensation for the children's possible neurological deficit. This project continues with a direct assessment of the significance of the Tallal's spectral manipulations for normal children and adults and will subsequently address the same issues in a population of LI children.

References

Lacerda, F. & Lindblom, B. (1998). Some Remarks on Tallal's Transform in the Light of Emergent Phonology. In C.von Euler, I. Lundberg, & R. Llinás (Eds.), *Basic Mechanisms in Cognition and Language* (1 ed., pp. 197-222). Amsterdam: Elsevier.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., & Tallal, P. (1996). Temporal Processing Deficits of Language-Learning Impaired Children Ameliorated by Training. *Science, 271,* 77-81.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., & Merzenich, M. M. (1996). Language Comprehension in Language-Learning Impaired Children Improved with Acoustically Modified Speech. *Science, 271,* 81-84.