

Acoustic analysis of Italian [r] and [l]

Francesca Argiolas, Federico Macrì, and Maria-Gabriella Di Benedetto

Dept. INFOCOM, Università degli Studi di Roma 'La Sapienza', via Eudossiana 18,
00184 Rome, Italy
e-mail: gaby@infocom.ing.uniroma1.it

paper presented at the 129th meeting of the Acoustical Society of America, Washington
D.C., June 1995.

Analysis method and reference patterns

Introduction

The present study aims at investigating spectral and temporal properties of [r] and [l] in Italian, in non-geminated and geminated forms. This investigation is aimed at the specification of formant reference patterns for these consonants.

The present study was based on the analysis of speech materials which were recorded for this specific purpose. The speech materials consisted of 'VCV Italian utterances spoken by 6 native Italian speakers (3 male and 3 female), in three repetitions. The vowel in the utterances was [i,a,u] and the consonant was [r,l] in non-geminated and geminated forms. The data-base consisted of 216 utterances.

Method

In the first phase of the study, an analysis of the speech materials was carried out in order to determine reference formant patterns for the consonant (formant frequency and temporal variations).

For this purpose, the following procedure was applied:

- the formant frequencies were measured by first applying an automatic algorithm, and secondly, by manual verification and adjustment of the values;
- the utterances were manually segmented and the segment durations were evaluated;
- the above data were statistically analyzed;
- the acoustic properties were perceptually validated.

Reference patterns for [r] and [l] in non-geminated and geminated forms

The diagrams presented show the F1, F2, F3, and F4 average trajectories, for each vowel and consonant (in geminated and non-geminated form).

Acoustic analysis

Time-domain analysis (1)

The duration of the first vowel in the VCV utterances was estimated. An average of the values for the geminated and the non-geminated consonants was computed. Results show that:

- the duration of the first vowel is systematically shortened in the geminated utterances;
- the duration of the first vowel is essentially independent of the vowel and consonant.

On the average, the ratio between the first vowel duration in geminated and non-geminated form is 70%.

Time-domain analysis (2)

The VCV utterances were analyzed in order to determine the durations of the VC and CV transitions in the geminated and non-geminated forms.

Results show that the VC and CV transition durations were not significantly different in geminated vs non-geminated forms.

Time-domain analysis (3)

The consonant duration was estimated. The values obtained were averaged over all speakers but the distinction with respect to vowel and consonant was maintained.

Results show that:

- the consonant is always longer in a geminated utterance than in a non-geminated one;
- the consonant duration is not significantly related to the consonant and vowel identity.

On the average, the ratio between non-geminated consonant duration and geminated one is 34%.

Time-domain analysis (4)

The duration of the second vowel in the VCV utterances was estimated. An average of the values obtained for geminated and non-geminated forms was computed. Results show that:

- the duration of the second vowel is systematically shortened in the geminated utterances;
- the duration of the second vowel is essentially independent of the vowel and consonant.

On the average, the ratio between the second vowel duration in geminated and non-geminated form is 87.5%.

Frequency-domain analysis (1)

The analysis in the frequency domain showed that there is no effect on the spectral properties of the vowel due to the presence of gemination and of a change in the consonant.

Similarly, the average formant frequencies of the consonants kept almost the same values in simple and geminated forms.

Average values \pm standard deviations of F1, F2, F3, and F0 frequency values of consonants [r] and [l] in the geminated and non-geminated form, for each vowel, are reported.

Frequency-domain analysis (2)

While the presence of a different consonant ([l] or [r]) has negligible effects on the formant frequencies of vowels, vowel identity influences the formant frequencies of the consonant.

It was shown previously that gemination does not influence consonant formant patterns. Consequently, formant patterns of [l] and [r] were obtained for each vowel and consonant by averaging the values obtained for the geminated and non-geminated forms. The figure shows the results obtained for the male speakers.

Comparative analysis

Spectral properties of [l] and [r]

- in some cases [r], mostly in geminated form, is devoiced. Two spectrograms of ['ar:a] are shown. In the first case, the [r] is voiced, while in the second case the [r] is devoiced.
- the spectrograms of [l] and [r] look similar in terms of energy amplitude and peaks location. The main difference between [l] and [r] spectra is the lack of occlusion in [l] for which the formants are straightly continuous.

Basis of the perceptual experiments

Perceptual experiments were performed in order to validate the remarks of the previous acoustic analysis. The synthesized utterances were obtained by patching different portions of speech signals in the time domain.

1st experiment: the close likeness of [l] and [r] spectrograms and short-time energy (with the exception of a short occlusion in [r]) suggest to substitute the stationary portion of [l] with a brief portion of silence.

2nd experiment: the role played by first vowel duration and consonant duration in the perception of gemination was investigated in the second experiment. Synthesized utterances were produced on the basis of the natural utterances of the data-base by simply modifying the duration of the first vowel and of the consonant according to the average values obtained in the time-domain analysis.

Perceptual experiments (1)

A synthesized [ara] was obtained by modification of a natural [ala] sample. The synthesized [ara] was obtained by substituting a silent portion of 12.8 msec to the

stationary part of [l]. The spectrogram on the left shows the synthesized [ara] sample. For comparison, the spectrogram on the right shows a natural [ara] of the same speaker.

The synthesized and natural [ara] show very similar spectral characteristics. These sentences were also very similar perceptually.

Perceptual experiments (2)

A synthesized geminated [r] in a [ar:a] utterance was obtained from a natural [ara] by repeating the consonant in the natural sentence three times, and by taking a the first vowel duration equal to 70% of the natural one.

The spectrograms show a comparison of a natural [ar:a] (pronounced by the same speaker of the original natural [ara]), and the synthesized [ar:a].

Although the natural [ar:a] and the synthesized [ar:a] are very similar perceptually, they exhibit, in terms of spectral properties, significant differences. The most striking difference, is that the synthesized [r] is voiced (since the original non-geminated [r] was voiceless) while the natural [r] is voiceless.

Perceptual experiments (2) (continued)

A synthesized [l] in a [ala] utterance was obtained from a natural [l:] of natural [al:a] by reducing the duration of the consonant to one third of the original value, and by modifying the duration of the first vowel.

The spectrograms show a comparison of a natural [ala] (pronounced by the same speaker of the original natural [al:a]), and the synthesized [ala].

In the present case, the spectrograms of natural and synthesized [l] are very similar. The consonants were also very similar perceptually.

In addition, informal perceptual tests, indicate that, in order to obtain a non-ambiguous non-geminated [l] from a geminated [l:], the first vowel duration must be increased. This observation is also valid for [r].

Interpretation

- it is impossible to give a gemination effect to a VCV utterance without lengthening the consonant (by about three times the duration of the single consonant), while it is possible to obtain such an effect without altering the first vowel duration (although results of informal perceptual tests indicate that this would result in a lack of naturalness). This result is in agreement with a previous study on gemination of Italian stops (R.Rossetti, 1994).
- acoustic and perceptual similarity of natural and synthesized [l:] suggest that gemination of [l] is a strengthening of the same phoneme (with different consonant duration).

- spectrographic differences between natural and synthesized [r:] (or [r]) suggest that [r:] is an independent long phoneme rather than a geminated cluster.
- acoustic and perceptual likeness of natural and synthesized [ara] confirms the importance of the short occlusion(s) (or energy amplitude decrease) in [r] (about 13 msec for each occlusion), as a sufficient but non necessary condition for the perception of a good [r].

References

F. Argiolas, 1995, *Analisi acustica e percettiva delle consonanti liquide [l,r] in italiano*, Laurea thesis, Univerisity of Rome 'La Sapienza', Italy.

F. Macrì, 1995, *Raddoppiamento nelle consonanti liquide [l,r]: acustica e percezione*, Laurea thesis, Univerisity of Rome 'La Sapienza', Italy.

R. Rossetti, 1994, "Gemination of Italian stops", *Journal of the Acoustical Society of America*, **95**, 2874.

INTRODUCTION

Goals:

- Spectral and temporal properties
 - Formant reference patterns
 - Acoustic correlates of gemination
- } of [r] and [l] Italian consonants.

Speech material

- Database of Italian natural utterances 'VCV' $\left\{ \begin{array}{l} V \in \{ [a], [i], [u] \} \text{ (Italian extreme vowels)} \\ C \in \{ [l], [r] \} \text{ (simple and geminated form)} \end{array} \right.$
- 216 utterances: 2 consonants \times 3 vowels \times (3 male speakers + 3 female speakers) \times 3 repetitions.

- spectrographic differences between natural and synthesized [r:] (or [r]) suggest that [r:] is an independent long phoneme rather than a geminated cluster.
- acoustic and perceptual likeness of natural and synthesized [ara] confirms the importance of the short occlusion(s) (or energy amplitude decrease) in [r] (about 13 msec for each occlusion), as a sufficient but non necessary condition for the perception of a good [r].

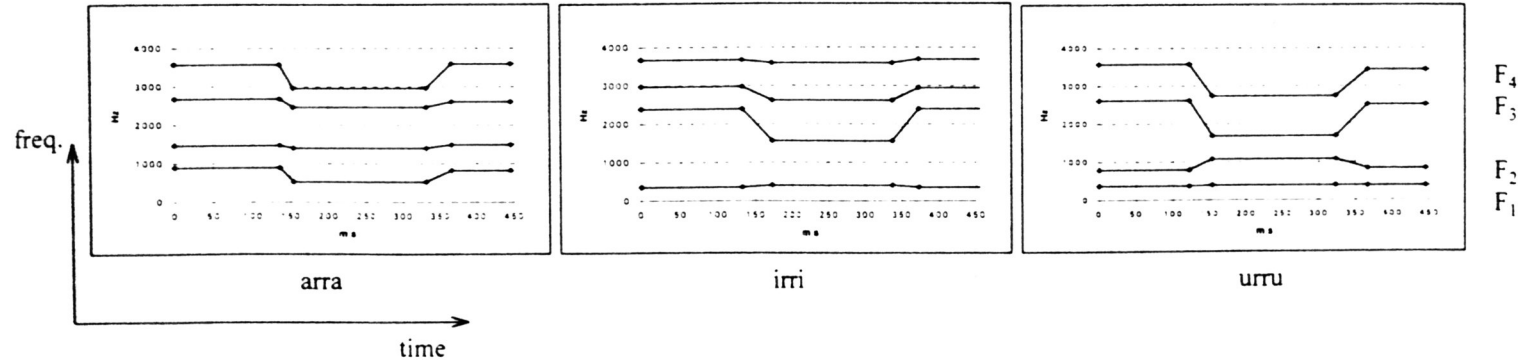
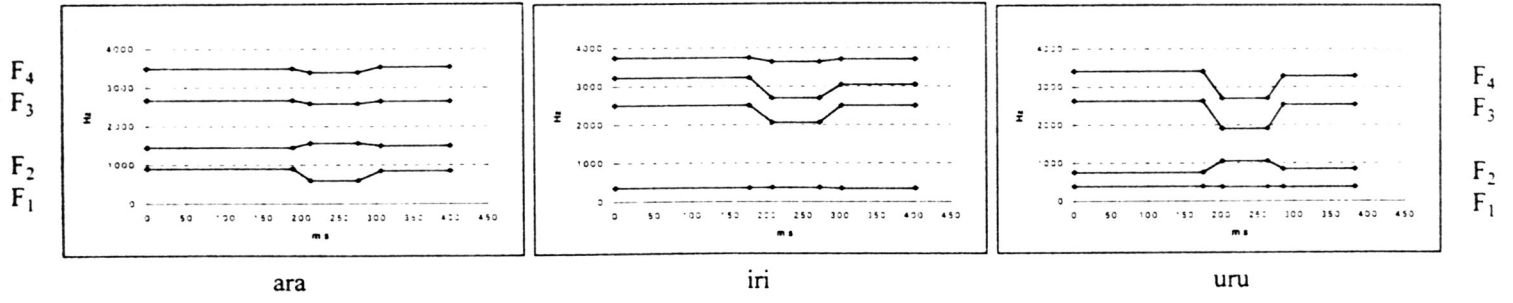
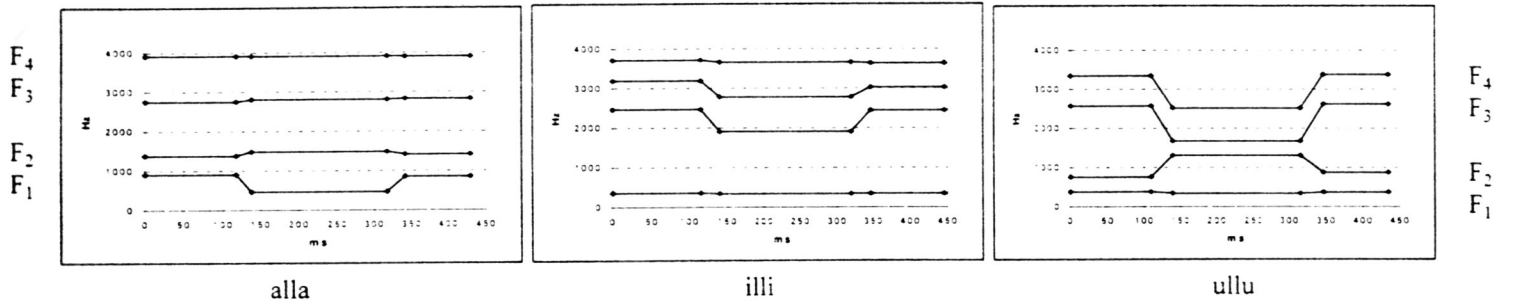
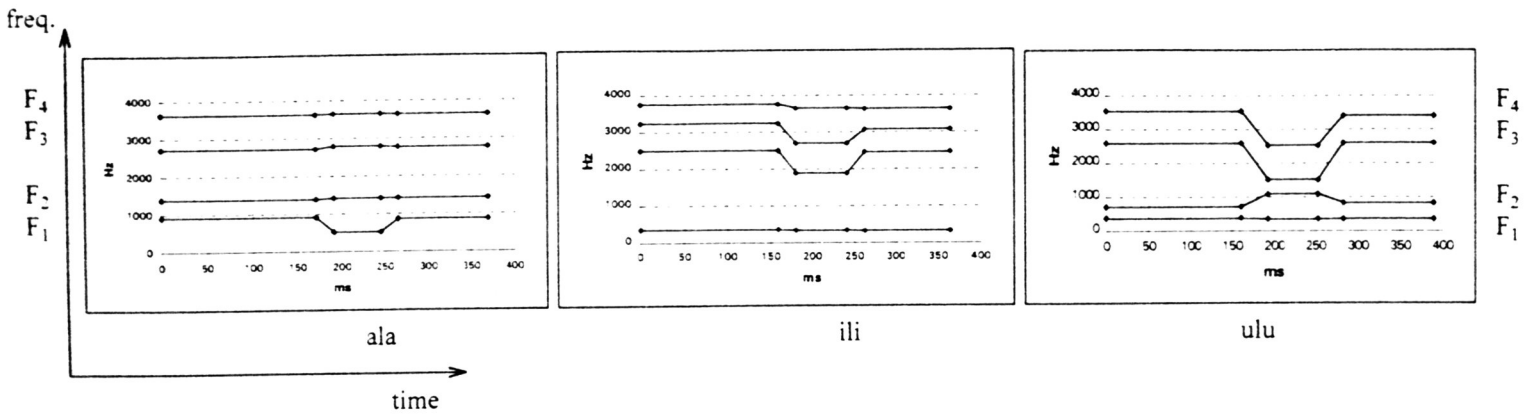
References

- F. Argiolas, 1995, *Analisi acustica e percettiva delle consonanti liquide [l:r] in italiano*, Laurea thesis, University of Rome 'La Sapienza', Italy.
- F. Macri, 1995, *Raddoppiamento nelle consonanti liquide [l:r]: acustica e percezione*, Laurea thesis, University of Rome 'La Sapienza', Italy.
- R. Rossetti, 1994, "Gemination of Italian stops", *Journal of the Acoustical Society of America*, **95**, 2874.

ANALYSIS METHOD

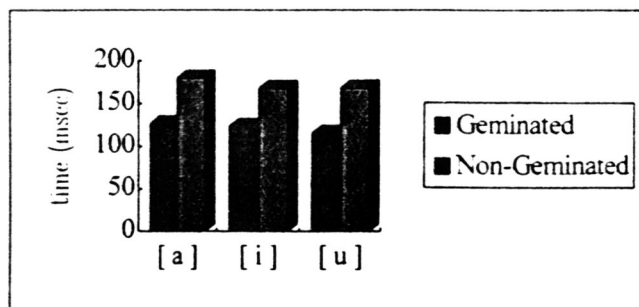
- 1) Formant frequencies measurement (automatic formant tracking followed by manual verification and correction)
- 2) Manual segmentation of utterances and segment duration measurements
- 3) Statistical analysis of the above data in time and frequency

The following reference patterns of a VCV utterance for [l] and [r], simple and geminated, coarticulated with [a], [i], [u] were obtained:

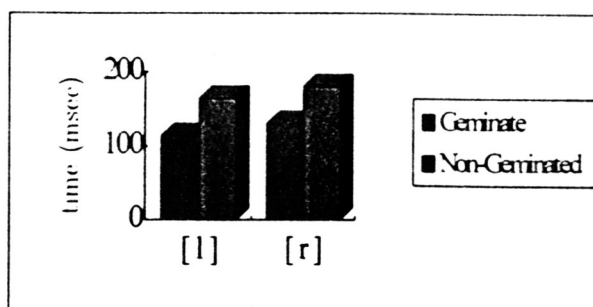


TIME - DOMAIN ANALYSIS (1)

First Vowel Duration (FVD) in a geminated vs. non-geminated context.



Average FVD in [aCa]-[aC:a], [iCi]-[iC:i], [uCu]-[uC:u] utterance, C ∈ {[l], [r]}.



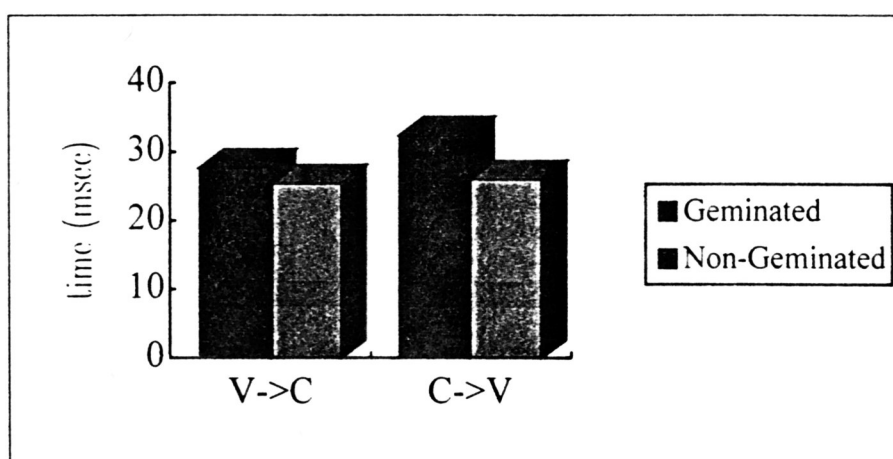
Average FVD in [VlV]-[Vl:V], [VrV]-[Vr:V], utterance, V ∈ {[a], [i], [u]}.

The previous diagrams show that the first vowel duration is systematically shortened in the geminated utterances. FVD is not significantly dependent of considered vowel and consonant.

$$\text{On the average: } \frac{FVD_{gem}}{FVD_{non-gem}} = 70\%$$

TIME-DOMAIN ANALYSIS (2)

V→C and C→V transitions duration in a geminated vs. non-geminated context.

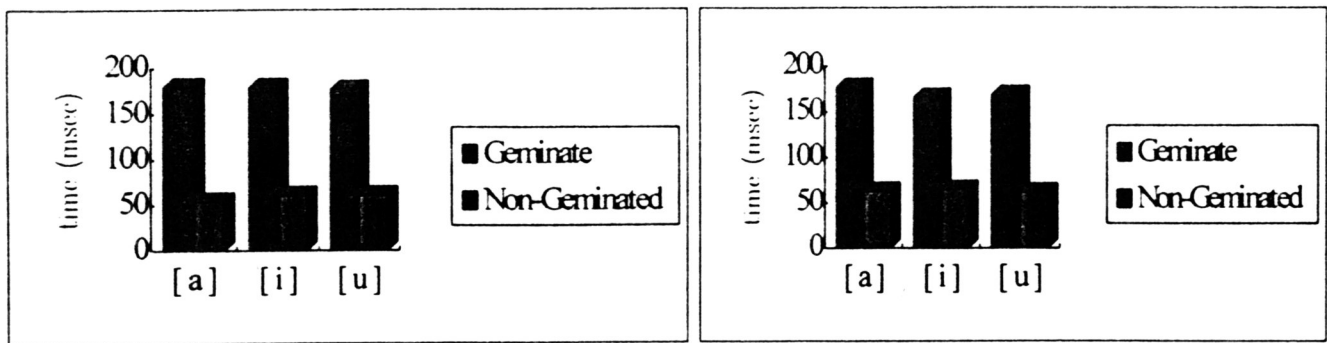


Average V→C and C→V duration.

V→C and C→V transitions have similar duration in geminated and non-geminated form.

TIME-DOMAIN ANALYSIS (3)

Consonant Duration (CD) in a geminated vs. non-geminated form.



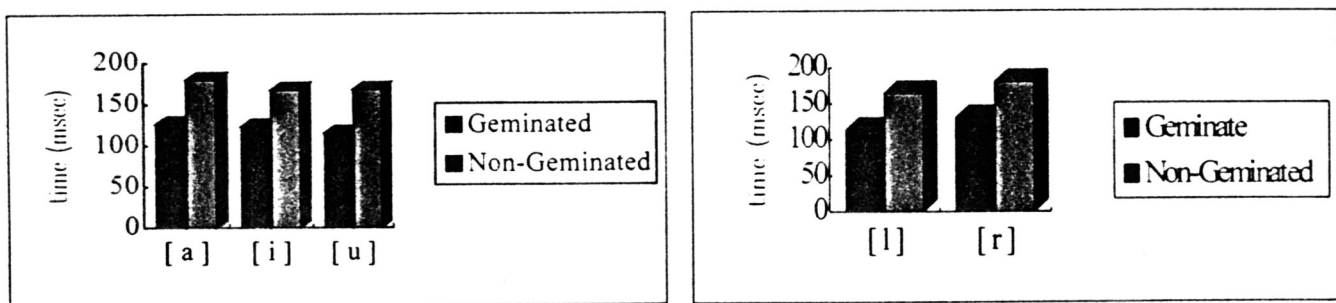
Average CD in [ala]-[al:a], [ili]-[il:i], [ulu]-[ul:u] utterances. Average CD in [ara]-[ar:a], [iri]-[ir:i], [uru]-[ur:u] utterances.

The previous diagrams show that the consonant duration is always longer in a geminated utterance than in a non-geminated one. CD is not significantly related to the considered consonant or coarticulated vowel.

$$\text{On the average: } \frac{CD_{non-gem}}{CD_{gem}} = 34\%$$

TIME-DOMAIN ANALYSIS (4)

Second Vowel Duration (SVD) in a geminated vs. non-geminated context



Average FVD in [aCa]-[aC:a], [iCi]-[iC:i], [uCu]-[uC:u] utterance, C ∈ {[l], [r]}.

Average FVD in [VlV]-[Vl:V], [VrV]-[Vr:V] utterance, V ∈ {[a], [i], [u]}.

The previous diagrams show that the second vowel duration is systematically shortened in the geminated form. SVD is not significantly dependent of considered vowel or coarticulated consonant.

$$\text{On the average: } \frac{SVD_{gem}}{SVD_{non-gem}} = 87.5\%$$

FREQUENCY-DOMAIN ANALYSIS (1)

- This analysis showed that there is no effect on the spectral properties of the vowel due to gemination and consonant identity.
- The average formant frequencies of geminated and non-geminated consonants are very similar.

	[a]		[i]		[u]	
	[l]	[l:]	[l]	[l:]	[l]	[l:]
F ₀	145±49	150±53	182±79	168±50	156±53	168±49
F ₁	518±65	471±43	338±59	344±62	351±60	344±52
F ₂	1417±274	1483±272	1895±173	1909±196	1102±122	1311±10
F ₃	2791±320	2811±340	2703±256	2779±237	1525±13	1678±62
F ₄	3651±364	3915±287	3657±303	3665±300	2542±20	2519±16

Mean values ± standard deviation of fundamental frequency and formants (F₁, F₂, F₃, F₄) for [l] in different vowel contexts. Values are in Hz.

	[a]		[i]		[u]	
	[r]	[r:]	[r]	[r:]	[r]	[r:]
F ₀	145±44	146±48	161±50	155±59	161±55	163±55
F ₁	600±69	527±31	367±48	409±56	379±48	394±32
F ₂	1565±227	1400±199	2061±297	1563±160	1054±46	1078±54
F ₃	2588±194	2468±221	2694±302	2618±277	1907±210	1684±155
F ₄	3392±187	2961±219	3642±480	3596±472	2701±298	2736±350

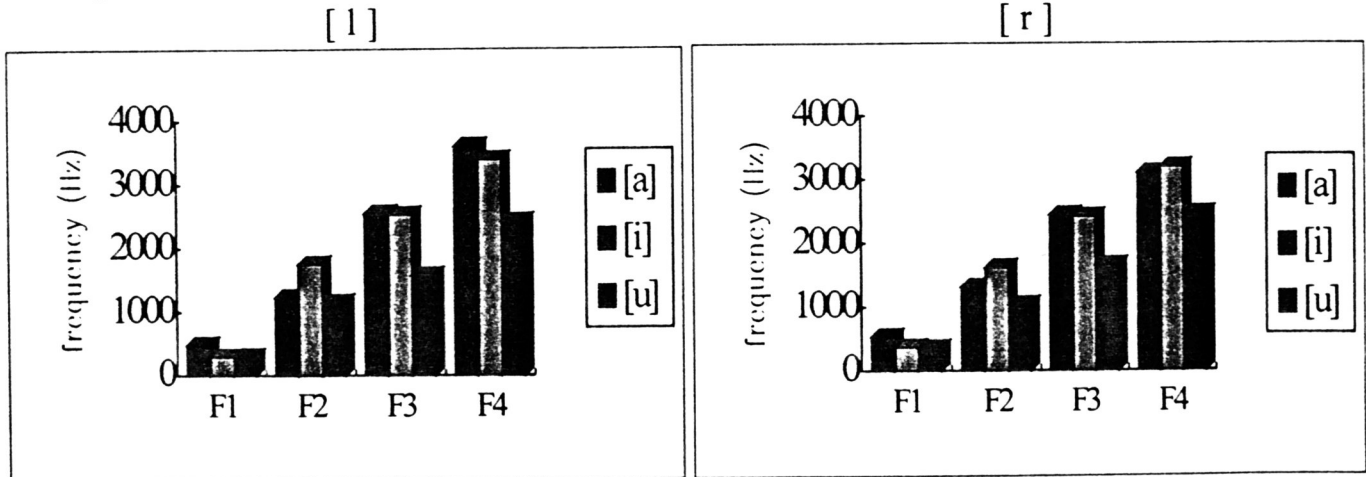
Mean values ± standard deviation of fundamental frequency and formants (F₁, F₂, F₃, F₄) for [r] in different vowel contexts. Values are in Hz.

14

FREQUENCY-DOMAIN ANALYSIS (2)

Dependence of consonant formant pattern on vowel identity.

Example: male speakers.

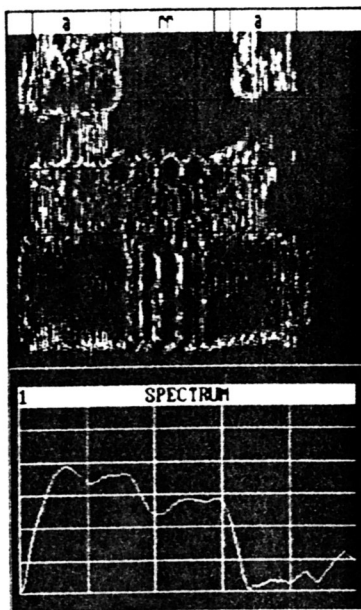


F₁-F₄ of [l] (geminated and non) depending on V ∈ {[a], [i], [u]}.

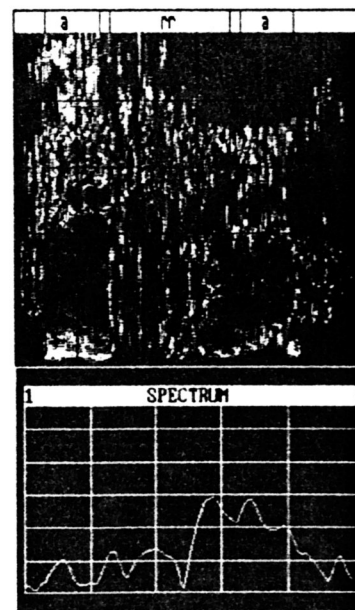
F₁-F₄ of [r] (geminated and non) depending on V ∈ {[a], [i], [u]}.

SPECTRAL PROPERTIES OF [l] AND [r]

- In some cases [r] (especially when geminated) is devoiced.



Spectrogram and wide-band spectrum of a voiced [ar:a]



Spectrogram and wide-band spectrum of a devoiced [ar:a]

- Spectrograms of [l] and [r] are similar in terms of energy amplitudes and peaks location. Difference between [l] and [r] spectra is the lack of occlusion in [l].

PERCEPTUAL EXPERIMENTS - SYNTHESIS TECHNIQUE

Perceptual experiments

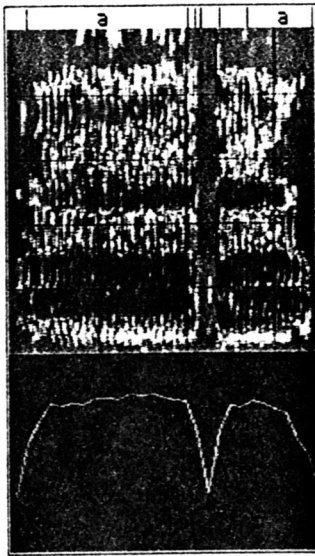
- Experiment 1: The close likeness of [l] and [r] spectrograms and short-time energy, with the exception of a short occlusion in [r], suggested to substitute the stationary portion of [l] with a brief portion of silence.
- Experiment 2: The role of FVD and CD in the perception of gemination was investigated. Synthesized utterances were obtained from the natural ones by modifying FVD and CD, according to the average values of the acoustic analysis results.

Synthesis technique

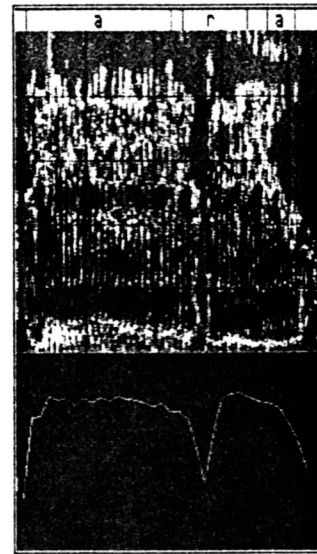
The synthesized utterances were obtained by patching different portions of the natural utterances in the time domain.

Perceptual experiment (1)

Comparison of a synthesized [ara] and a natural [ara].



Synthetic [ara]



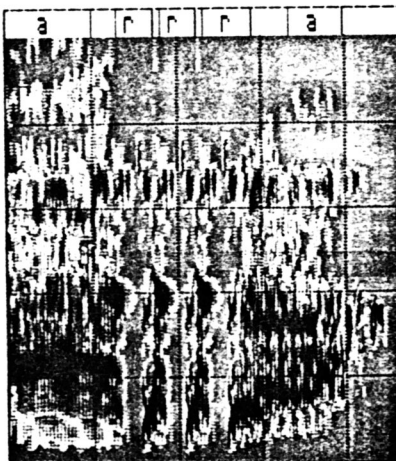
Natural [ara]

No substantial difference between artificial and natural utterances in terms of spectral properties.

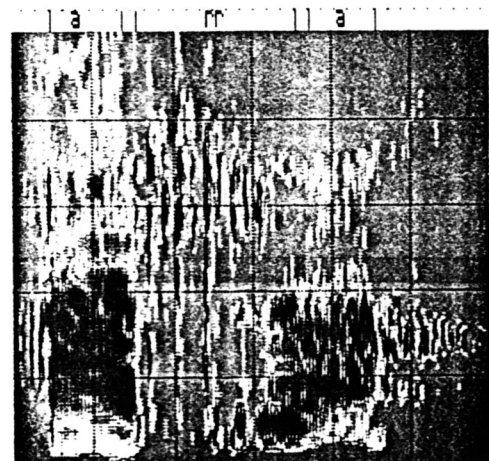
The synthesized natural and [ara] were also perceptually very similar.

Perceptual experiment (2)

Comparison of a synthesized [ar:a] and a natural [ara].



Synthetic [ar:a]

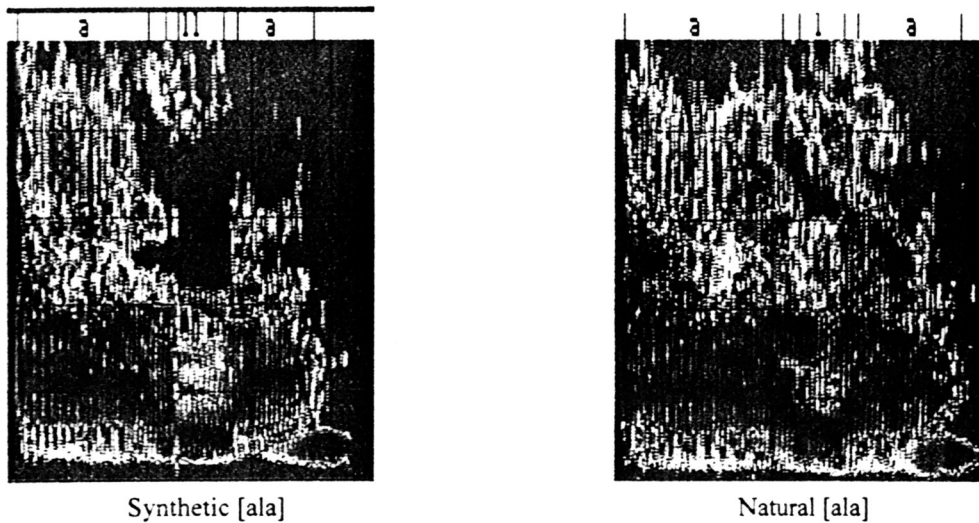


Natural [ar:a]

The natural and synthesized [ar:a] are perceptually very similar. However, they are different in terms of spectral properties (synthesized is voiced while natural is voiceless).

Perceptual experiment (2) (continued)

Comparison of a synthesized [ala] and a natural [ala].



The synthetic and natural [ala] are very similar in terms of spectral and perceptual properties.

INTERPRETATION

- It is impossible to give a gemination effect to a VCV utterance without lengthening the consonant (by about 3 times the duration of a single consonant), while it is possible to obtain such an effect without altering the FVD (although results of informal perceptual tests indicate that this would result in a lack of naturalness).
- Spectrographic differences between natural and synthesized [Vr:V] (or [VrV]) suggest that [r:] is an independent long phoneme rather than a geminate cluster.
- Acoustic and perceptual likeness of natural and synthesized [ara] confirms the importance of the short occlusion(s) (or energy amplitude decrease) in the [r] (about 13 msec for each occlusion) as a sufficient but non necessary condition for the perception of a good [r].

REFERENCES

- Argiolas Francesca, 1995, *Analisi acustica e percettiva delle consonanti liquide [l, r] in italiano*; Laurea thesis in Electronic Engineering, University of Rome “La Sapienza”.
- Macri Federico, 1995 *Raddoppiamento nelle consonanti liquide [l, r]: acustica e percezione*; Laurea thesis in Electronic Engineering, University of Rome “La Sapienza”.
- Rossetti Riccardo, 1994, “Gemination of Italian stops”, *Journal of the Acoustical Society of America*, **95**, 2874.