VOICING ASSIMILATION IN CATALAN AND ENGLISH

by

Néstor Cuartero Torres

TESI DOCTORAL

Universitat Autonòma de Barcelona

Departament de Filologia Anglesa i de Germanística

2001

ACKNOWLEDGEMENTS

I would like to express my gratitude to the speakers who volunteered for the experiment and to my colleagues at the Universitat Autònoma de Barcelona. I also want to thank the staff at the Phonetics Lab of the Universitat de Barcelona.

I would like to thank my supervisor, Maria Josep Solé for her wise advice, encouragement and proofreading of earlier versions of this dissertation.

I could not have completed this work without the help and encouragement of my parents and the rest of my family. I would like to thank David for his patient proofreading.

Last but not least, I want to dedicate this work to Monica and Mariona, for their unconditional love.

TABLE OF CONTENTS

Page

1. Introduction 1

1.1. Main features of glottal and supraglottal activity 5

1.1.1. Phonation: voicing and voicelessness 5

1.2. Coarticulation versus assimilation 9

1.3. Voicing agreement in obstruent sequences 12

1.3.1. Voicing agreement across languages 13

1.3.2. Voicing assimilation in Catalan 17

1.3.2.1. Empirical findings on Catalan sequences 20

1.3.3. Voicing agreement in English 28

1.3.3.1. Empirical findings on English sequences 31

Relevance and aims of the study 40

2. Method 42

2.1. Material 44

2.2. Subjects 49

2.3. Experimental procedure 50

2.3.1. Electropalatography 51

2.3.2. Electroglottography 54

2.4. Analysis procedure 56

Summary 62

3. Results 63

3.0. Preliminary remarks 65

3.1. Voiced – voiceless stop sequences 68

    3.1.1. Speaking rate effects 74

    3.1.2. Degree of overlap 76

3.2. Voiceless – voiced stop sequences 81

    3.2.1. Speaking rate effects 85

    3.2.2. Degree of overlap 88

3.3. Fricative – stop sequences 91

    3.3.1. Voiced – voiceless sequences 91

    3.3.2. Voiceless – voiced sequences 96

    3.3.3. Speaking rate effects 100

    3.3.4. Summary 103

3.4. Sequences of nasal and obstruent consonants 104

    3.4.1. Obstruent – nasal sequences 105

    3.4.2. Nasal – obstruent sequences 112

    3.4.3. Speaking rate effects 113

    3.4.4. Effect of degree of overlap 114

3.5. Sequences of obstruent and lateral consonants 119

    3.5.1. Obstruent – lateral sequences 120

    3.5.2. Lateral – obstruent sequences 122

    3.5.3. Speaking rate effects 127

    3.5.4. Effect of degree of overlap 128

Summary 134

4. Discussion and Conclusions 136

4.1. Discussion 138

4.2. Phonological modeling 144

4.3. Conclusions 157

REFERENCES 158

APPENDIX A: Catalan and English data 171

Catalan data 171

English data 193

APPENDIX B: Stimuli 212

APPENDIX C: Sample data 216

1. Introduction

1. Introduction

One of the most important goals of phonetic theory is to formulate an explicit account of the detailed phonetic features of human languages, and to explain their relationship to the linguistic units that underlie their production. That understanding of this relationship is central to phonetics is evidenced by the research that explores the interface between phonetics and phonology.

The present study intends to analyze the process of voicing assimilation across different speaking rates in Catalan and English. More specifically, it aims at observing and characterizing voicing assimilation when two consonants that have a different phonological specification for voicing co-occur across word boundaries. Furthermore, this study intends to evaluate current descriptive frameworks, in order to see how they account for the data obtained in the experiment.

It has been observed that the process of voicing assimilation has a different direction and extent in Catalan and English (Cuartero 1998). This suggests that phonological units with the same feature specification are phonetically implemented differently in both languages. Thus, a sequence like Catalan *pot dur* /∩π t# δυ/ results in complete regressive assimilation of the voicing feature, [∩π d# δυ]. Conversely, in the English sequence *back door* /∩βΘκ# δ ù/ C1 remains voiceless and C2 is partially or fully devoiced, [∩βΘκ# d8 ụ]. The present study aims at examining such cross-linguistic differences in the phonetic implementation of phonological units with the same feature specification using original data, and to review how current phonetic and phonological theories model such differences.

We also intended to analyze the influence of speaking rate and articulatory overlap on voicing assimilation in the two languages. Various studies (Barry 1985, Kerswill 1985) have shown that, amongst other factors, the overall tempo of performance affects assimilatory processes and connected speech processes in general. More concretely, there seems to be a trend towards more assimilation in fast speaking rates, and less assimilation in slow rates of speech. It has also been shown that assimilation is a gradient process (*e.g.* Nolan & Holst 1993) which may co-vary with speaking rate. Thus, we intended to test whether the patterns of co-ordination of oral and glottal gestures are categorical or gradient across different speaking rates in each of the languages under study.

As Solé (1995a, 1995b) argues, observing the effect of speaking rate on assimilatory processes may help determine the automatic or phonological nature of the phenomenon. Thus, if extent and magnitude of assimilation does not vary with speaking rate it may be due to the fact that it is a phonological phenomenon, whereas if it does vary with speaking rate it may be due to the fact that it is a phonetic process.

Additionally, it has been claimed (Holst and Nolan 1995) that assimilatory phenomena may stem from the same source (articulatory overlap) but may reflect two different processes. At one end of the range, gestural overlap may result in varying degrees of articulatory assimilation. At the other end, these coarticulatory tendencies may have been encoded in a categorical phonological process giving rise to complete assimilation.

In sum, research on voicing assimilatory processes has proved to be interesting for descriptive purposes and leads to a better understanding of connected speech processes. A fine-grained analysis of the coordination of oral and glottal gestures in consonant sequences may throw light on the nature of voicing assimilation in Catalan and English and has implications for linguistic theory. Finally, the understanding of this phenomenon is of interest for educational and technological applications.

This chapter is divided into four main sections. In section 1.1 the main features of glottal and supraglottal activity in obstruents and sonorants are discussed. Section 1.2 deals with the concepts of assimilation and coarticulation. In section 1.3 the process of voicing assimilation in Catalan and English is described.

1.1. Main features of glottal and supraglottal activity

In this section, the main features of glottal gestures will be described. Emphasis will be put on those characteristics of speech production which are important for the description of voicing in Catalan and English obstruents and sonorants.

1.1.1. Phonation: voicing and voicelessness

Phonation is the use of the laryngeal system to produce an audible source of energy that can be modified by the articulatory actions of the vocal apparatus. This section deals with the different ways the larynx can produce voicing and voicelessness. The larynx is capable of producing a wide range of different modes of phonation (Laver 1994: 187), but here we will only concentrate on those aspects of laryngeal activity that are relevant for the description of voicing in obstruents and sonorants.

Vibration of the vocal folds in voicing creates a pulsed input, and the frequency of the pulsing is the product of muscular and aerodynamic factors. Vocal fold vibration can be described on the basis of the aerodynamic-myoelastic model of phonation (Müller 1837, van den Berg 1958, 1962), which explains how air-pressure and airflow factors on the one hand, and mechanical factors in laryngeal muscles on the other, contribute to the production of voicing. The explanation that follows, then, is based on this model.

For voicing to take place, the adductor muscles of the larynx have to close the larynx, so that the inner edges of the vocal cords are in light contact. Pressure from the pulmonic airstream builds up and subglottal pressure rises. As a result, the vocal cords are blown slightly apart and the compressed air below the glottis flows through this narrow gap in the larynx at a very high speed due to high subglottal pressure. Then, as a result of the Bernoulli effect (a pressure drop caused by acceleration of the air through a very narrow constriction), there is a decrease in air pressure along the lateral walls of the vocal folds that forces the reestablishment of the glottal closure. The tension of the laryngeal muscles also contributes to the closure of the glottis. This process is repeated many times at a very high speed, and is conditioned by many factors, such as the presence of sufficient pressure difference between the oropharyngeal and subglottal cavities. The acoustic correlate of voicing is a periodic wave that corresponds to the opening and closing gestures of the vocal folds.

Whereas the maintenance of voicing in sonorant consonants is not problematic because the air escapes freely through the oral or through the nasal cavities, the maintenance of glottal vibration is in particular danger in obstruents due to the fact that the exit of air from the oral cavity is partially or fully blocked, which leads to a rapid buildup of supraglottal air pressure, and to the cessation of transglottal airflow and eventually of voicing.

Stop consonants are produced by a complete closure in the vocal tract. The complete oral closure in oral stops is combined with a velic closure that prevents the air from escaping through the nasal cavity. As a result, there is a rise in intra-oral pressure, so that, when the closure is released, the compressed air escapes to the atmosphere with a stop burst and pressure falls rapidly. Voiceless oral plosives have a volume of air of approximately 80 ml (Warren 1996). In voiced oral stops, the volume of air used is approximately 50 ml, and air is impounded for about 125 ms. The pressure magnitude, between 3 and 5 cm H_2O (Warren 1996), is lower for voiced stops, due to glottal impedance. In addition, voiced stops are always shorter than voiceless stops. If voiced stops were as long as voiceless ones, voicing would die out since subglottal and oral pressure would equalize. For voicing to take place the vocal folds have to be in contact and there has to be sufficient pressure drop across the glottis. In long stops and fricatives air pressure in the oropharyngeal cavity builds up as air form the lungs accumulates behind the stop/fricative constriction. With time, oral pressure and subglottal pressure equalize and transglottal flow, and, consequently, voicing stops.

Fricatives are produced with a stricture of close approximation. The cross-sectional area of the constricted aperture has to be small enough to cause audible friction (Laver 1994: 244). It is also important to consider aerodynamic factors to explain the production of fricative sounds. Voiceless fricatives are the consonants that involve the greatest respiratory effort. Pressure magnitude ranges from 3 to 8 cm H_2O to ensure that air from the lungs flows through the narrow constriction at a high velocity (Warren 1996), causing air turbulence. Voiceless fricatives have a volume of air of about 100 ml and they have a longer duration than voiced fricatives (Smith *et al*. 1978). Voiced fricatives, on the other hand, involve less respiratory effort. Also, they involve a smaller volume of air (75 ml. approximately) and have a shorter duration (about 125 ms). Voiceless fricatives are more frequent in the world's languages than their voiced counterparts (Maddieson 1984, Maddieson and Ladefoged 1996). According to Maddieson and Ladefoged (1996: 176-178), this is due to acoustic as well as aerodynamic reasons. Firstly, the strong low-frequency energy resulting from vocal fold vibration masks the lower-amplitude frication in the higher frequency range. Secondly, the flow impedance at the glottis makes it difficult to create turbulence at the articulatory constriction.

Voicing in obstruents can be maintained if pressure buildup is delayed and this may be achieved by expanding the cavity behind the closure. For example, it can be achieved through pharyngeal expansion – by lowering the larynx or puffing up the cheeks – which allows more air to accumulate in the pharyngeal cavity, so that the time during which there is sufficient pressure drop across the glottis is extended. Thus, if the pharynx is expanded, voicing will be favored during an obstruent. Other maneuvers to prolong the continuation of vocal fold vibration during a stop or a fricative constriction include relaxation of the soft tissues around the oropharyngeal cavity, so that the pressure will passively expand the volume, moving the articulatory constriction forwards during the closure, moving the tongue root forwards or lowering the jaw (Bell Berti 1975, Ohala and Riordan 1979, Keating 1984). Since maintenance of vocal fold vibration is dependent on oropharyngeal cavity size, degree of voicing during obstruent constriction varies with place of articulation of the consonant: sounds whose place of articulation is near the front of the mouth provide larger surfaces of soft tissue in the vocal tract walls, the yielding of which permits more air to be accommodated supralaryngeally before voicing dies out.

According to Catford (1977), laryngeal activity is of a gradual nature. Thus, laryngeal gestures form a continuum that ranges from absolute voicelessness with vocal fold abduction to complete vocal fold adduction as in a glottal stop, and voicing would be in-between these two poles. This implies that transition from voicing to voicelessness is not instantaneous. Voicing tends to begin and end gradually and, for example, in the transition from voicelessness to voicing there is a phase where the vocal cords vibrate without coming together thus producing a short period of breathy voice.

In this dissertation, however, the continuum of possible modes of phonation will be disregarded, because the main objective is to analyze the timing of voicing with respect to supraglottal articulators, rather than to study the exact shape of glottal movements. Speech sounds will be identified as either voiced or voiceless, as has been done in other studies of voicing in obstruents (*e.g.* Docherty 1992).

1.2. Coarticulation versus assimilation

The autonomy of the segment has been often questioned. It is well known that speakers do not articulate segments in isolation in connected speech. Rather, segments influence each other, a phenomenon that has been referred to as *coarticulation*. Adjacent segments affect each other in running speech in such a way that their articulations interact and overlap in both anticipatory and perseverative fashion. In other words, the articulation of segments is sensitive to context.

The term ‘coarticulation’ was first used by Menzerath and Lacerda (1933). This phenomenon has also been referred to as ‘articulatory smoothing’ (Fujimura and Lovins 1978: 108) and ‘similitude’ (Jones 1960). As Clark and Yallop point out (1990: 120), this term is not used consistently in the literature. Many writers use the term to refer indistinctly to all kinds of assimilation processes, no matter how many articulators are involved or the nature of the process itself. In the present dissertation, a clear distinction will be made between coarticulation on the one hand and assimilation on the other, since they are processes of a different nature (Recasens 1993).

On the one hand, coarticulation is a purely phonetic phenomenon that takes place in all human languages and involves articulatory adaptation of a segment to its phonetic environment without altering its phonological features. Moreover, coarticulatory phenomena are obligatory and occur automatically (Recasens 1993: 54). For instance, in VC sequences where C is a velar stop, the oral velar gesture of the consonant starts during the preceding vowel. At the same time, the front/back nature of the vowel will determine an advanced or retracted articulation of the velar consonant. Similarly, in English words like *can* and *pan* the velum is lowered during the vowel preceding the nasal stop, in anticipation of the following segment. In brief, coarticulation is always present, although it can vary depending on factors like speaking rate (Amerman, et al. 1970) and the language being spoken (Boyce 1990).

Assimilation, on the other hand, has a phonological nature, since it involves a different input to the articulatory program, that is, the modification of the phonological features of a segment due to the influence of an adjacent segment. Moreover, assimilatory processes may or may not be present, depending on factors such as speech rate and speaking style (Recasens 1993: 53-54). Assimilation occurs, for example, in English alveolar consonants preceding an interdental fricative (*e.g.*, *ten things* /∩τεν TINζ/ → [∩τHεν5 TI)Nζ6]). Assimilation may also be present in the Catalan sequence *pot bo* (/∩π τ β / → [∩π β β] 'good pot'), where the alveolar stop in word final position assimilates the bilabial and voicing feature of the following stop.

If we look more closely at these sequences, however, important differences emerge. In *pot bo*, the sequence of stops is heterorganic, that is, the consonants in the sequence /t#b/ involve articulators that can move independently from each other: the tongue tip for the alveolar constriction for /t/, and the lips for the labial constriction for /b/. Thus, the articulators involved in the production of the consonant sequence can be manipulated independently of each other. One implication of this is that the bilabial gesture for C2 may be anticipated, so that it may overlap the alveolar gesture for C1.

In the sequence *ten things*, on the other hand, the two adjacent consonantal phonemes /νT/ constitute a *contiguous* sequence (Catford 1977). Both require adjacent parts of the same articulator for their production (the tongue tip and blade) with different targets (apico-alveolar and lamino-dental). Thus, both segments have conflicting targets for *the same* articulator. According to Browman and Goldstein (1990), the motor commands for the two adjacent segments overlap in running speech and the articulator involved in the articulatory conflict reaches a compromise with the articulatory demands of the adjacent interdental fricative. Solé and Estebas (1995), however, found that cases of competing demands on the same articulator for consonants involving equally constrained gestures were solved in favor of C2. In any case, this second phenomenon is best described as accommodation. In brief, accommodation can be viewed as 'a particular case of the more general phenomenon known as assimilation' (Catford 1977: 224). The process of voicing assimilation in consonant sequences, by which a given segment acquires the voicing features of a neighboring sound, is one example of accommodation since it involves the overlap of conflicting motor commands directed to the same articulator, the vocal folds.

It is important to point out that coarticulation and assimilation are two related phenomena. It is the case that the perceptual reinterpretation of coarticulatory effects is responsible for assimilation (Ohala 1990, Recasens 1999). Assimilation occurs when listeners are unable to compensate for regressive or progressive coarticulation. Thus, assimilation has both a coarticulatory and perceptual basis.

1.3. Voicing agreement in obstruent sequences

As noted above, the overlap of motor commands in running speech results in competing demands on the glottis for sequences of consonants that have a different phonological specification for voicing. Thus, in the Catalan two-word sequence *cap gran* ('big head'), there is a conflict in articulatory demands, since the word-final stop requires glottal abduction and the word-initial stop requires vibration of the vocal folds. Exactly the same situation occurs, but word-medially, in the compound *capgros* ('tadpole'), where the bilabial stop is in morpheme-final position. The outcome of this articulatory conflict in Catalan is that the bilabial stop accommodates its glottal articulation to that of the following stop (*i.e.* [∩κɑβ gPan], [∩κɑβ gP ɑ]). Thus, two successive different states of the glottis –abducted and adducted – are replaced by a single state that is maintained unchanged. This phenomenon has often been described as Regressive Voicing Assimilation.

The ensuing subsections deal with the process of voicing agreement in obstruents. Section 1.3.1 looks at voicing agreement across languages, whereas sections 1.3.2 and 1.3.3 focus on the process of voicing assimilation in Catalan and English. Subsections 1.3.2.1 and 1.3.3.1 deal with empirical findings on Catalan and English, respectively. It must be noted that studies of voicing assimilation in obstruents across word boundaries are scant in the literature, and that the situation is still worse with regard to phonetic experiments in this field. As G.J. Docherty (1992: 41) puts it, 'there has been scarcely any controlled instrumental investigation of this type of coordinatory activity in English consonant sequences'. This is applies to other languages, including Catalan.

1.3.1. Voicing agreement across languages

As is well known, there is a cross-linguistic tendency among the world's languages by which obstruent clusters agree in voicing. In fact, generative phonology (see, for example, Mester and Itô 1989) has interpreted the process as a phonotactic constraint of Universal Grammar. In Ancient Greek, for example, the voicing specification of the initial obstruent in the suffix replaces the voicing specification of the final consonant in the stem, as shown in (1). The inflectional suffixes add a past/perfective meaning to the stem.

(1) / τριβ ù/ 'rub' / τετριπ + ται/
/ κλεπτ ù/ 'steal' / κλεβ + δΕù ν/

In Yiddish, there is also assimilation to the voicing of the last obstruent in a cluster, as exemplified in (2):

(2) (examples from Lombardi 1999)
a) [ϖογ] 'weight'
[ϖοκΣοι] 'scale'
b) [βριϖ] 'letter'
[βριφτρεγερ] 'postman'
c) [ζισ] 'sweet'
[ζιζϖαργ] 'candy products'

Ukrainian shows voicing assimilation when the second consonant in the sequence is voiced, as illustrated in (3a). However, there is no assimilation when C2 is voiceless, as shown in (3b):

(3)

a)	*molo*[τ]	‘to mill’
	molo[δβ]*a*	‘milling’
b)	*ri*[δκ]*o*	‘rare’
	ve[ζτ]*y*	‘to drive’

Most of the examples given above correspond to cases where there is voicing agreement (except in 3b) in word-internal obstruent clusters. This process also occurs frequently across word-boundaries. In the example from Spanish in (4), the word-final obstruent acquires the voicing specification of the initial consonant in the following word:

(4)	*los*	[λοσ]	‘the’
	los buenos	[λοζ Βωενοσ]	‘the good ones’
	los cortos	[los koPtos]	‘the short ones’

Hungarian also has a rule of voicing assimilation whereby obstruent clusters come to share the voiced/voiceless specification of their rightmost member across word boundaries as well as within words, as illustrated in (5)

(5) *Ismered a fiút, boldog* ‘You know the boy, he is happy.’
[fiju:d boldog] [t] → [d]
A víz hideg ‘The water is cold.’
[vi:s hideg] [z] → [s] (example from Gósy 1999)

In Syrian Arabic, regressive voicing assimilation occurs across word boundaries. An example of voicing is given in (6a) and an example of devoicing is given in (6b) below:

(6) (example from Barry 1999)

a) /Σtareùt daùr/ → [Σtareùd daùr] ‘I bought a house’

b) /lwalad taùni/ → [lwalat taùni] ‘another child’

As the examples in (1) to (6) show, voicing agreement is most frequently a regressive process and this seems to be a universal tendency. However, there are some languages that show progressive assimilation of voicing. One typical example is voicing agreement in English inflectional suffixes, which must agree in voicing with the last consonant of the stem. The example in (7) below shows that the phonological specification for the final consonant spreads rightward to the plural suffix /Iζ/ after the deletion of the intervening vowel (Kenstowicz 1994).

(7) a) *cat* /κΘτ/ *cats* /κΘτ/ + /Iz/ → [kHΘτσ]

b) *dog* /δ γ/ *dogs* /δ γ/+ /Iζ/ → [δ8 γζ§

In Swedish, a voiced obstruent always devoices next to a voiceless obstruent. One consequence of this is that this language displays a bi-directional spread of the feature [-voice]. The examples in (8) illustrate this phenomenon.

(8) (examples from Lombardi 1999)

a) <u>Regressive assimilation of voicelessness</u>

hög	[ηöù γ	‘high’
högtid	[η↓κù τιù ϕ	‘festival’

b) <u>Progressive assimilation of voicelessness</u>

the preterite underlying dental suffix *-de* devoices to *–te*.

/stek/+/de/ → /stekte/ ‘fried’

In sum, voicing assimilation across words reflects the overwhelming tendency to maintain the same glottal configuration for continuous segments. The following sections analyze in more detail the situation in Catalan and English.

1.3.2. Voicing assimilation in Catalan

In all Catalan dialects, obstruents assimilate the voicing feature of an immediately following obstruent both within and across word boundaries. This process seems to be part of the speaker's phonological system. Thus, in consonant sequences where the first is underlyingly voiceless and the second is underlyingly voiced, the former is phonetically implemented as a fully voiced consonant (Badia i Margarit 1994, Recasens 1993, Carbonell 1992). The examples in (9) below illustrate this phenomenon.

(9) Regressive assimilation of voicing

a) within words

capgros 'tadpole', noun /∩καπ γP σ/ →
[∩καβ γP σ]

pasdoble *'paso doble',* noun /∩pασ δοβλ↔/ →
[∩pαζ Δοπλ↔]

b) across word boundaries

nap gran 'big turnip' /∩ναπ# γPαν/ →
[∩ναβ γPαν]

gos danès *'Danish dog'* /∩γοσ δ↔ νEσ/ → [∩γοζ
Δ↔ νEσ]

Voicing assimilation does not occur in Catalan if there is a major pause or a major boundary between the two consonants.

In the same way, when a voiceless obstruent follows an underlyingly voiced consonant, there is leftward spreading of the [-voice] feature to the first consonant in the sequence. This is illustrated in (10):

(10) Regressive assimilation of voicelessness

set	‘thirst’	/σΕδ/
set terrible	‘terrible thirst’	/ σΕδ#τ↔ ριπλ↔/→ [σΕτ τ↔ ριπλ↔]
cas	‘case’	/καζ/
cas terrible	‘terrible case’	/ καζ#τ↔ ριπλ↔ /→ [καστ↔ ριπλ↔]

In *set*, the underlyingly voiced stop is devoiced due to a process of Final Obstruent Devoicing (FOD). Thus, in Catalan, there is neutralization of voicing in final obstruents, which may lead to ambiguities; for example, the noun *set* (/σΕδ/ ‘thirst’) in (10) is homophonous with the numeral *set* (/σΕτ/ ‘seven’), as shown in (11):

(11)

set terrible	‘terrible thirst’	/ σΕδ#τ↔ ρι bλ↔/→[σΕτ τ↔ ρι pλ↔]
set terribles	‘seven terrible ones’	/ σΕτ#τ↔ ρι bλ↔ s/ →[σΕτ τ↔ ρι pλ↔σ]

In the same line, word-internal obstruent sequences must have the same voicing specification. Some examples are provided in (12):

(12)

aptitud	‘aptitude’	[↔πτι τυτ]
abdicar	‘to abdicate’	[↔βδι κα]

astorar	‘to frighten’	[↔στυ ρα]
esvalotar	‘to agitate’	[↔ζΒ↔λυ τα]

Lexical items like *obtenir* (‘to obtain’) seem to be apparent exceptions to this tendency. However, these words show regressive assimilation of voicing, *i.e.* [υπτ↔ νι]. It may well be possible that, though voicing assimilation is not reflected in the spelling, it is present in the speakers’ phonology. In fact, spelling mistakes like **dissapte*, for *dissabte* (‘Saturday’), seem to suggest that this is the case. Thus, words like *futbol* ([φυδ β λ], ‘soccer’), *obtús* ([υπ τυσ], ‘obtuse’), and *obtenir* ([υπτ↔ νι], ‘to obtain’) cannot be considered counterexamples to the general rule of voicing agreement in obstruent sequences in Catalan.

In sum, the overlap of conflicting consecutive glottal gestures in Catalan obstruent sequences is resolved by anticipating the glottal gesture of C2 to the first consonant in the sequence. The situation is summarized in Table I below.

Table I. Voicing assimilation in Catalan obstruent sequences within and across word boundaries.

Process	Sound-type	Context	Example
FOD	All obstruents	___#	tub /τυβ/→ [τυπ]
Voicing Assimilation C1 > voiced	All obstruents	___C $_{[+voice]}$	*ca*/π/-*gros* →*ca*[β]-*gros*
	All obstruents	___#C$_{[+voice]}$	*na*/π/ *dolç* → *na*[β] *dolç*
Voicing Assimilation C1 > voiceless	All obstruents	___C$_{[-voice]}$	*obtús* [υπ τυσ]
	All obstruents	___#C$_{[-voice]}$	*Se*[δ] *terrible* → *se*[τ] *terrible*

1.3.2.1. Empirical findings on Catalan sequences

Dinnsen and Charles-Luce (1984) set up an experiment in order to see whether there is any phonetic difference between word-final voiced and voiceless stops when the following word starts with a voiceless consonant (*e.g.*, *ce**c** **c**larament* ‘blind clearly’ /γ#κ/, *se**c** **c**larament* ‘dry clearly’ /κ#κ/). They measured three variables from acoustic data: 1) duration of the vowel preceding C1, 2) Voicing into the closure of C1 and 3) closure duration of C1.

Dinnsen and Charles-Luce (1984) found no difference between final voiceless and voiced stops in the group results for any of the three variables. However, they did find individual differences. Firstly, it was seen that, for one speaker, the duration of vowels was longer when they preceded underlyingly voiced stops. Secondly, they report that the duration of the closure of underlyingly voiced stops was clearly shorter for one of the subjects. Their conclusion was that the strategies for maintaining the voicing distinction could vary across speakers.

The experiment outlined above had a number of flaws (*see* Mascaró 1987 for details). The most serious ones were the inadequate choice of lexical items and the assumptions made about their underlying forms. For example, they considered that the Catalan words *cap* ‘head’ and *cap* ‘it fits’ both had an underlying final voiced stop. In addition, neither the stress pattern differences nor the grammatical status of the lexical items used in the experiment were controlled.

In Charles-Luce and Dinnsen (1987) the data were reanalyzed after correcting the errors concerning the choice of underlying forms of the lexical items used in the experiment. Thus, in this second experiment, they properly considered that the underlying form of the Catalan verb *cap* ‘it fits’ was /καβ/. They measured the same variables as in the previous experiment and they did not consider grammatical status or stress pattern.

In the group results, Charles-Luce and Dinnsen (1987) did not find significant differences between final voiced and voiceless stops for any of the three variables. Concerning the individual results, they observed that there was a trend toward a voicing distinction in one or more of the variables studied, although they did not reach statistical significance because of the small number of items studied. The general conclusion of Dinnsen and Charles-Luce's experiment was that voicing differences are phonetically maintained, even if this is not perceptually salient for the native speakers of that language.

In spite of the methodological problems of these experiments, the question of neutralization of voicing distinctions in obstruent sequences is still interesting and needs further empirical support. Concerning the sequences that are the object of this dissertation, the question is raised whether in Catalan sequences like *pot gran*, where C1 completely assimilates the voicing feature of C2, the voicing distinction is still present in some other way.

Carbonell (1992) claims that voicing assimilation is a deep-rooted process that forms an integral part of the sound system of Catalan, and he does so on the basis of three facts. Firstly, native speakers of Catalan transfer the rule of voicing assimilation to any language they learn, even when this second language does not show Regressive Voicing Assimilation. For instance, Catalan learners of English would pronounce the English sequence *sit down* with a voiced C1 [∩σIδ δαΥν]. Secondly, misspellings involving voicing assimilation occur very often. For example, the words *aràcnid* 'arachnid' and *dissabte* 'Saturday' are often spelt *aràgnid* and *dissapte*. Thirdly, borrowings adapted from English are affected by the voicing assimilation rule, such as the word *futbol* ‘soccer’, which is pronounced [φυδ βολ] or [φυb βολ].

In his experiment, Carbonell (1992) analyzed acoustic and laryngographic data from three native speakers of Catalan. As in the two experiments discussed above, he measured three variables – duration of the vowel that preceded C1, duration of the closure of C1, and degree of voicing in the closure of C1 – in the context $\mathrm{C1}_{[\text{-/+voice}]}\#\mathrm{C2}_{[\text{+/-voice}]}$ in order to see whether neutralization of the voicing distinction was complete.

In the group results, no significant differences were found for underlying voicing in any of the variables in the context $\mathrm{C1}\#\mathrm{C2}_{[\text{+voice}]}$. (*see* Figure 1.1). In the individual results, no significant differences between voiced and voiceless stops were found for vowel duration or closure duration. As for voicing into the closure, it was found that there was more voicing in the consonant, either voiced or voiceless, when it was followed by a voiced, as opposed to a voiceless stop. Thus, there was voicing assimilation. Two of the speakers did not show any difference between underlyingly voiced and voiceless stops in C1 position. One of the speakers, however, presented significantly more voicing into the closure of voiced plosives compared to voiceless plosives (*see* Figure 1.2) in the context ___$\#\mathrm{C2}_{[\text{+voice}]}$.

According to Carbonell (1992), the results described above provide evidence that there exists a rule of Final Voicing Assimilation. Voiced and voiceless plosives have more voicing when followed by voiced stops than when followed by voiceless stops both in the individual and in the group results. However, his results also suggested that neutralization of the voicing distinction is incomplete, since the data from one of the speakers showed that underlyingly voiced stops had significantly more voicing than their voiceless counterparts when followed by a voiced stop.

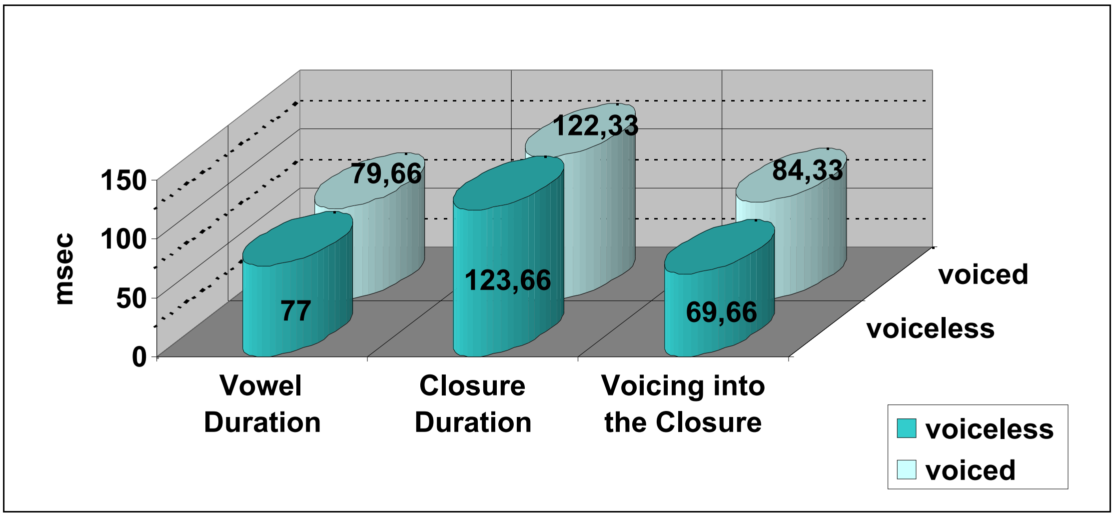

Figure 1. 1. Group results for the three variables, on the abscissa, studied in Carbonell (1992). The figure shows that there are no significant differences between underlyingly voiced and voiceless stops for any of the variables in the context ___$C2_{[+voice]}$.

In sum, Carbonell reached the same conclusion as Dinnsen and Charles-Luce (1984, 1987), namely that neutralization is subject to speaker variation. Thus, some speakers show complete neutralization; that is, a sequence of obstruents like /t#g/ is phonetically like /d#g/, whereas others seem to exhibit a voicing distinction in the phonetic realization of these sequences.

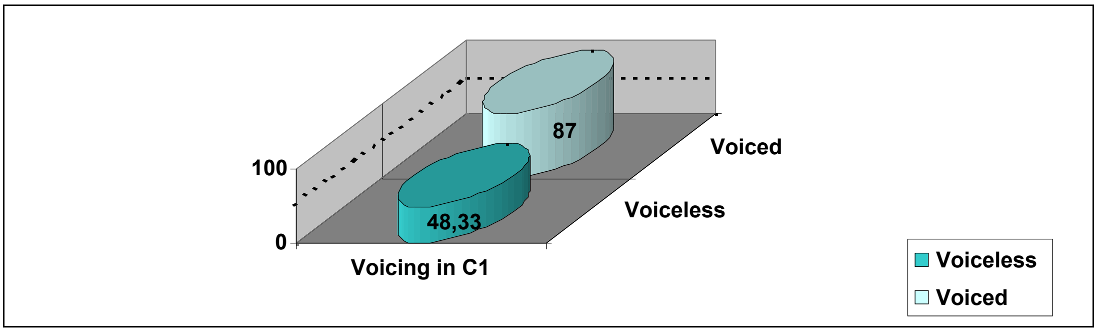

Figure 1. 2. Degree of voicing in C1 for speaker SJ in the context $C1_{[+/-voiced]}\#C2_{[+voiced]}$ (data from (Carbonell 1992). It can be seen that underlying voiced stops (light blue) show a greater degree of voicing into the closure than the voiceless stops (dark blue).

One important finding of Carbonell (1992) was that there was variability between speakers in the voicing patterns of consonant sequences where obstruents preceded nasal consonants. It was found that although some speakers assimilated voicing before a nasal, others did not show significant voicing. On the basis of these results, Carbonell (1992) concluded that voicing assimilation in Catalan is only triggered by a following obstruent. Carbonell claimed that there is a rule of regressive voicing assimilation in Catalan, but its application before a nasal is subject to speaker variation.

One important problem with these three experiments is that only the acoustic waveform was used to measure the duration of the closure of C1. Dinnsen and Charles-Luce (1984, 1987) considered that the first perturbation of energy in the waveform was the end (or release) of C1. These measurements are problematic for two reasons. Firstly, a stop can be released without causing a perceptible perturbation in the acoustic waveform. Secondly, in sequences of consonants uttered in casual speech, C1 is most often unreleased due to articulatory overlap. Therefore, analyzing only those sequences where C1 is released involves giving an unreal picture of the language being studied. In fact, Charles-Luce and Dinnsen (1987) acknowledge that 36 per cent of the sequences could not be analyzed because C1 was unreleased. Carbonell (1992) faced the same problem: he could not measure closure duration in C1 in most sequences because it was often unreleased, a problem that can be solved by using electropalatography.

Cuartero (1998) carried out an experiment with two native speakers of Catalan - labeled MJ and DA –and two native speakers of English –labeled ME and AL –comparing the co-ordination of oral and glottal gestures in voiceless – voiced stop sequences using EPG, laryngographic and acoustic data. The aim of the experiment was to observe the direction and the extent of voicing, and also to see whether voicing assimilation co-varied with speaking rate. As far as Catalan is concerned, the voicing timing patterns found in the mixed sequences (*i.e.,* voiced – voiceless and voiceless –

voiced sequences), together with the control voiced-voiced and voiceless – voiceless stop sequences are schematically presented in Figure 1.3 below:

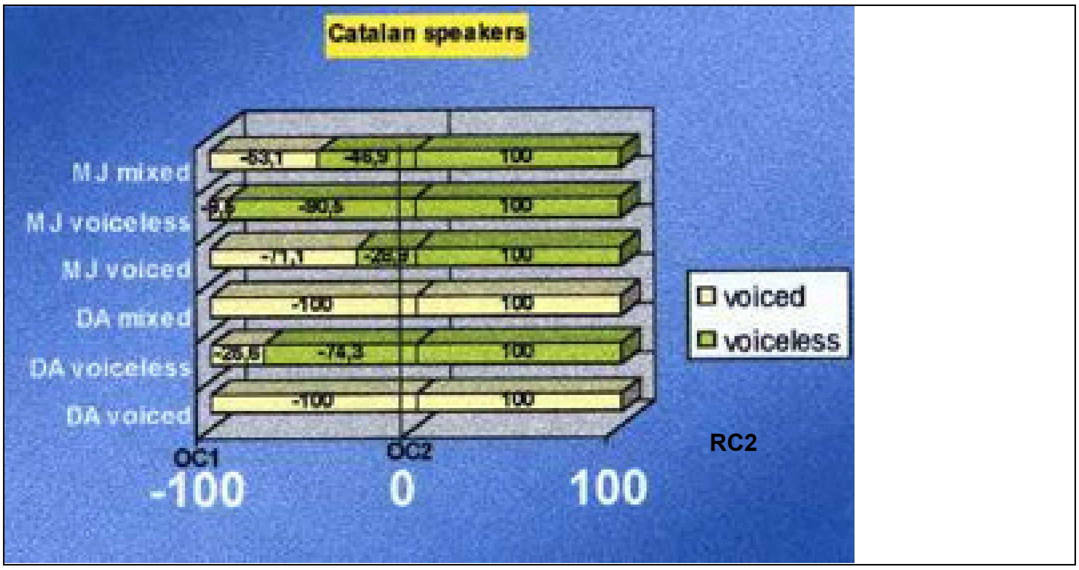

Figure 1.3. Percentage of voicing and voicelessness in the Catalan mixed and control sequences for speakers MJ and DA. (Cuartero 1998). OC1= onset of C1 closure; OC2 = onset of C2 closure; RC2 = release of C2 closure.

Figure 1.3 shows that in the voiced control sequences – MJ voiced and DA voiced in the Figure - C1 was voiced for both speakers, though one of the speakers (MJ) showed mechanical devoicing of the latter portion of C1 and C2. The voiceless control sequences were produced with voicelessness throughout the closure, although there was voicing continuation at the beginning of the closure of C1. In the mixed sequences, C1 was voiced for speaker DA, who exhibited no difference in voicing patterns between mixed and voiced sequences. Similarly, speaker MJ showed voicing in approximately the first half of C1, and mechanical devoicing during the second half of C1, which continued into C2. Thus, it was considered that the whole sequence had been planned as voiced. In sum, the Catalan speakers exhibited complete regressive voicing assimilation of voicing into C1 with variability between speakers in the maintenance of vocal fold vibration.

As mentioned above, Cuartero (1998) also investigated the effect of speaking rate on voicing assimilation in Catalan. The results showed no co-variation of regressive voicing assimilation with speaking rate. Sequences of stops where C1 was phonologically voiceless and C2 phonologically voiced were produced with a voiced C1, regardless of rate. Voicing assimilation, thus, applied categorically, and there were no cases of partial spreading of voicing.

In the line of the premises of Articulatory Phonology, Romero (1999) interprets voicing assimilation in Spanish as gestural blending. He carried out an experiment using EMMA data where he studied the effects at the gestural level of voicing assimilation in /s/ +voiced or voiceless dental, velar and alveolar stops within and across word boundaries. The experiment also analyzed the timing of oral and laryngeal gestures in these consonants in isolation.

Concerning the magnitude of the laryngeal gestures, Romero (1999) showed that single voiced stops have the highest levels of voicing; in other words, they have the smallest glottal opening. Voiceless consonant clusters, on the other hand, have the smallest level of voicing, that is, they have the greatest glottal opening. Romero (1999) also found that degree of voicing in /s/ - voiced stop sequences is always lower than in single voiceless consonants.

According to Romero (1999) these findings may imply either that Spanish voiceless stops are not fully voiceless in running speech or that there is no assimilation to the expected degree in /s/ - voiced stop sequences. In the second interpretation of the data, /s/ - voiced stop sequences would show laryngeal gestures whose magnitude is in-between those of a single voiced stop and those of a /s/ - voiceless stop sequence. Thus, voicing assimilation would not be a categorical phenomenon in Spanish.

As for the timing of laryngeal gestures with respect to supralaryngeal gestures, Romero (1999) showed that, in single consonants, the supralaryngeal peak occurs before the laryngeal peak (in voiceless consonants) or slightly afterwards (in voiced consonants), whereas in consonant clusters the laryngeal peak always precedes the supralaryngeal peak, due to the presence of the /s/. As Romero points out, the lag between supralaryngeal and laryngeal peaks in /s/ in the clusters is longer than the lag found in single voiceless stops, which shows that the laryngeal gesture is not synchronized with the supralaryngeal gesture for /s/, but occurs somewhere between the supralaryngeal peaks of /s/ and the following consonant in the sequence.

Thus, Romero (1999) concludes that the experiment challenges traditional accounts of voicing assimilation, which would predict a categorical spreading of the laryngeal configuration of C2 to C1. What the results of this study show is, firstly, that voicing assimilation in Spanish is gradual, that is, the magnitude of laryngeal gestures varies depending on the context. Secondly, the timing relationship between laryngeal and supralaryngeal gestures in /s/ - C clusters show mutual influence that results in blending of laryngeal gestures. The facts observed in this experiment clearly fit within the postulations of Articulatory Phonology, which predicts blending of same-tier articulatory gestures in running speech.

1.3.3. Voicing agreement in English

In most English dialects, sequences of obstruents differing in voicing show a different behavior from Catalan as regards not only the direction but also the extent of voicing assimilation. In obstruent sequences where a voiceless obstruent is followed by a voiced one, C1 is kept voiceless whereas C2 tends to become *partially* devoiced. This process is referred to as Progressive Voicing Assimilation, since the abductory gesture of the vocal folds for C1 is perseveratory and extends to the immediately following consonant. This occurs both within words and across word boundaries, as illustrated in (13a) and (13b) below:

(13) a) Within words

outbid (verb) /∩αυτ βΙδ/ → [∩αυτ β8Ιδ8]

casebook (noun) / κεΙσβΥκ/ → [κΗεΙσβ8Υκ]

c) Across words

hot beer (Noun + Adj.) /∩η τ# βΙ↔/ → [∩η τ
β8Ι↔]

this beer (Det. + Noun) /∩ΔΙσ βΙ↔/→ [∩Δ8Ισ β8Ι↔]

In sequences where a voiced obstruent is followed by a voiceless one, C1 becomes partially devoiced, so that the assimilatory process is regressive. This occurs both within and across word boundaries as illustrated in (14a) and (14b) below. The situation is summarized in Table II below.

(14) a) <u>Within words</u>

bagpipe (noun) / bΘγπαΙπ/ → [b8Θγ8παΙπ]

newspaper (noun) / νϕυù ζπεΙπ↔/ →

[νϕυù ζ8πεΙπ↔]

b) <u>Across words</u>

bad time (Adj. + Noun) /∩βΘδ ταΙμ/ →

[∩β8Θδ 8τΗαΙμ]

news conference (Noun+Noun) /∩νϕυù ζ κ νφ↔ρ↔νσ/→ [∩νϕυù ζ8

κΗ Μφ↔ρ↔νσ]

Complete regressive assimilation of voicelessness is found in sequences such as *have to* [ηΘφτυ], where the fricative takes the voicing specification of the following obstruent. Another example is the sequence *five pence*: /φαΙϖ/ + /πενσ/ → [φαΙφπενs]. However, this process only occurs in very common sequences that form a close-knit unit (*i.e.*, almost a lexical unit, Gimson 1970). In sum, complete voicing assimilation takes place in some special cases and progressive assimilation is the norm (Gimson 1970, Abercrombie 1967).

Complete regressive assimilation of voicing is found among speakers of Educated Scots (Abercrombie 1967) both within and across word boundaries, as shown in (15). Thus, speakers of Scottish English produce a voiced stop cluster in words like *blackboard*, as opposed to the majority of English dialects, which would produce C2 as a partially voiced stop and C1 as a voiceless stop.

(15) <u>Voicing assimilation in Educated Scots</u> (Abercrombie 1967)

- within words: *blackboard* [βλΘγβ ù]δ

- across words: *with them* [ωΙΔΔ↔μ]

Table II. Voicing assimilation in obstruent sequences in Standard English.

Process	Sound-type	Context	Example
Progressive C2 > partially devoiced	Obstruents	C[-voice]___	*Hot-**d**og* [η τδ8 γ]
	Obstruents	C[-voice]#___	*Hot **b**eer* [∩η τ β8l↔]
Regressive C1 > partially devoiced	Obstruents	___C[-voice]	*bagpipe* [ɓ8Θγ8παlπ]
	Obstruents	___#C[-voice]	*bad time* [∩β8Θδ 8τHαlμ]

Summing up, the phonetic implementation of voicing in sequences of obstruents varies cross-linguistically and cross-dialectally. The present study will be devoted to the analysis of voicing assimilation across word boundaries in sequences of consonants with different phonological specifications for voicing.

1.3.3.1. Empirical findings on English sequences.

Studies on voicing in English obstruents mainly focus on stop consonants, whereas voicing in fricatives and affricates has been investigated much less. According to Docherty (1992) most descriptions of the co-ordination of oral and glottal gestures in voiceless – voiced stop sequences in English show that speakers use more than one voicing timing pattern. This seems to parallel the variability found in the production of voiced stops in initial position, as schematically represented in Figure 1.4 below:

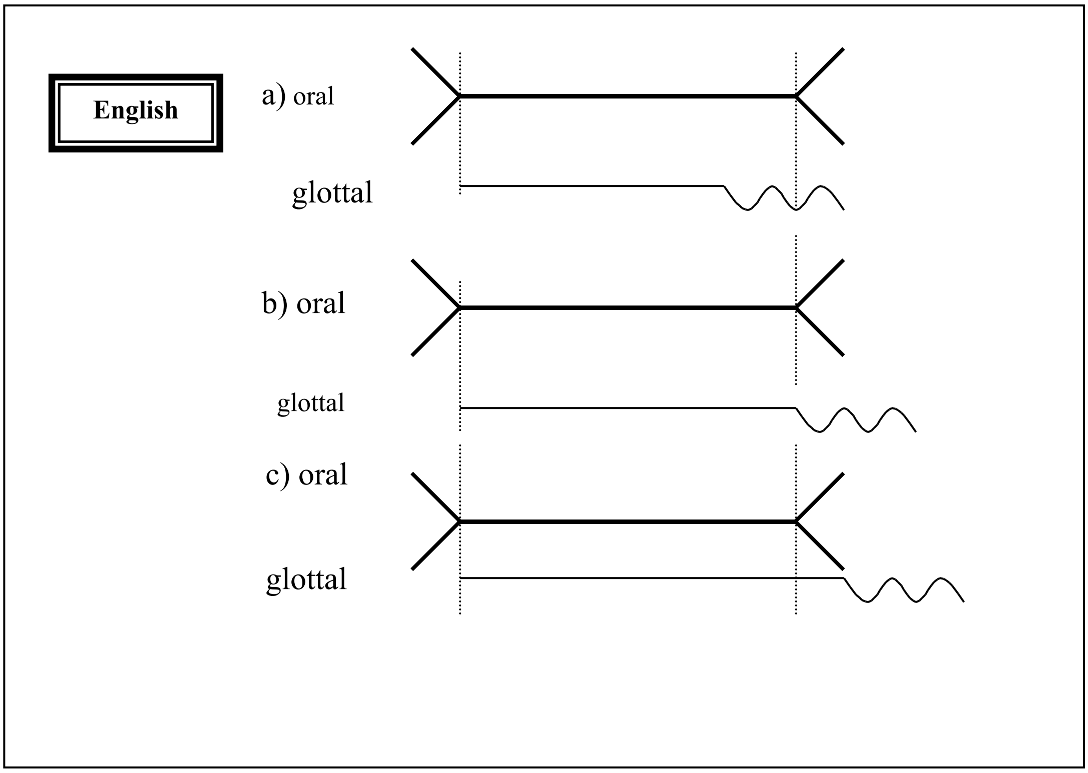

Figure 1. 4. Schematic representation of the oral (top) and glottal (bottom) articulations of initial voiced stops in English, based on Lisker and Abramson (1964).

In fact, there is disagreement among researchers as to which pattern is more frequently used. According to Lisker and Abramson (1964), short voicing lead – as in (a) in the figure – is the most common pattern, but Smith (1978) and Westbury (1979) did not confirm these results. Their experiments showed that there were no subjects who showed one pattern exclusively, although individuals usually presented a preference for using one pattern more than the other. In any case, it has yet to be seen how these patterns of coordination are affected when consecutive stops are uttered in running speech.

Docherty (1992) carried out two experiments using acoustic and laryngographic data in order to investigate the effect of underlying phonological category and phonetic context on voicing in stops and fricatives, and in sequences where stops and fricatives combine with other consonants in Standard British English. Docherty studied two parameters: VOT and voicing during the medial phase of obstruents.

Firstly, the experiments showed that voiced stops have significantly shorter VOT values than voiceless stops. For voiced stops, it was observed that speakers used two voicing timing patterns, one which involved commencing voicing before the stop release, and one where voicing started at the same time as or shortly after the stop release. Docherty (1992) points out that there is a lot of variability in VOT values within and across speakers, which contrasts with the claim by Stevens & Klatt (1974: 653) that VOT values of voiced and voiceless stops rarely overlap.

As for voicing during the medial phase of post-vocalic word-initial stops, *e.g.,* 'a pin, a bin', it was found that there is usually voicing continuation in both voiced and voiceless stops. Voicing continuation was very short when the stop was voiceless, whereas the percentage of voicing during the medial phase of voiced stops was higher than in voiceless stops.

In word-final stops, Docherty (1992) found that, in most cases, both voiced and voiceless stops had voicing continuation into the medial phase when there was a voiced sound preceding them. Complete voicing was only found in some occurrences of voiced stops, and complete voicelessness was found in underlyingly voiceless stops, although there were some occurrences of word-final voiced stops which were unvoiced.

In the few experiments carried out on sequences of voiceless – voiced consonants across word-boundaries, the most frequently noted pattern of co-ordination is late onset of voicing in C2. Thus, Jones (1960) observed that in consonants like /m, n, l/ there is a delay in voice onset for the consonant when preceded by a voiceless segment. Similarly, Docherty (1992) found more devoicing of C2 in voiced stop-sonorant sequences than in voiceless stop – sonorant sequences.

Westbury (1979) studied the timing of voicing in stop sequences in American English. More specifically, his experiment focused on word-medial sequences of stops. He used ten subjects, who produced a list of two-syllable words of the form ‘CVCCV(C). He found that there was a lot of variation between speakers concerning the voice timing patterns used, although he never found cases of regressive assimilation of voicing in voiceless – voiced sequences, *i.e.*, cases where voicing started during the closure of C1. The results were explained by referring to the voicing timing patterns observed in the production of single stops in initial position: subjects who produced voiced stops with voicing lead also showed voicing lead in sequences of voiceless – voiced stops. Subjects who produced voiced stops with short lag, on the other hand, showed the same pattern in stop sequences. Thus, he concluded that there were two patterns of co-ordination of glottal and oral gestures in voiceless – voiced stop sequences, which were analogous to the two patterns observed in single post-pausal voiced stops. In voiced – voiceless sequences, Westbury (1979) found that most speakers showed partial devoicing of C1. In some cases, it was also found that C1 was fully voiced.

One question none of these studies has addressed is how the co-ordination of oral and glottal gestures is affected by speaking rate cross-linguistically. That is, one wonders whether the process of regressive or progressive assimilation of voicing is gradual and co-varying with speech rate. If Barry and Kerswill's (1985) results on place assimilation apply to these sequences, that is, more assimilation the faster the speaking rate, then it is to be expected that the process will be gradual and co-varying with speaking rate in English. Nolan (1992) provides further evidence supporting the gradient nature of assimilatory processes. He showed that place assimilation in ALVEOLAR – VELAR sequences of stops is a gradual process.

Cuartero (1998) investigated the direction and extent of voicing in English voiceless – voiced stop sequences across word boundaries, as well as the influence of speaking rate on the co-ordination of oral and glottal gestures in these sequences. The patterns in the control voiceless – voiceless and voiced – voiced sequences together with the patterns in the mixed sequences /t#g/ and /k#d/ are schematically represented in Figure 1.5. In the voiced control sequences – AL voiced and ME voiced in the figure – C1 was fully voiced and C2 was also voiced though voicing tended to die out in the latter portion of the consonant constriction. The voiceless control sequences – AL voiceless and ME voiceless in the figure – presented absence of vocal fold vibration in both C1 and C2. As regards the mixed voiceless – voiced sequences, C1 was fully voiceless in all the utterances, so that there were no cases of regressive voicing assimilation. As for the production of C2, the subjects ME and AL exhibited two markedly distinct patterns, as shown in Figure 1.5. Speaker ME, on the one hand, showed complete devoicing of C2 in most cases, although there were three cases of partial devoicing. On the other hand, speaker AL exhibited a voiced C2 closure in most cases and mechanical devoicing during the latter portion of the constriction, although there were also some cases of partial devoicing of C2.

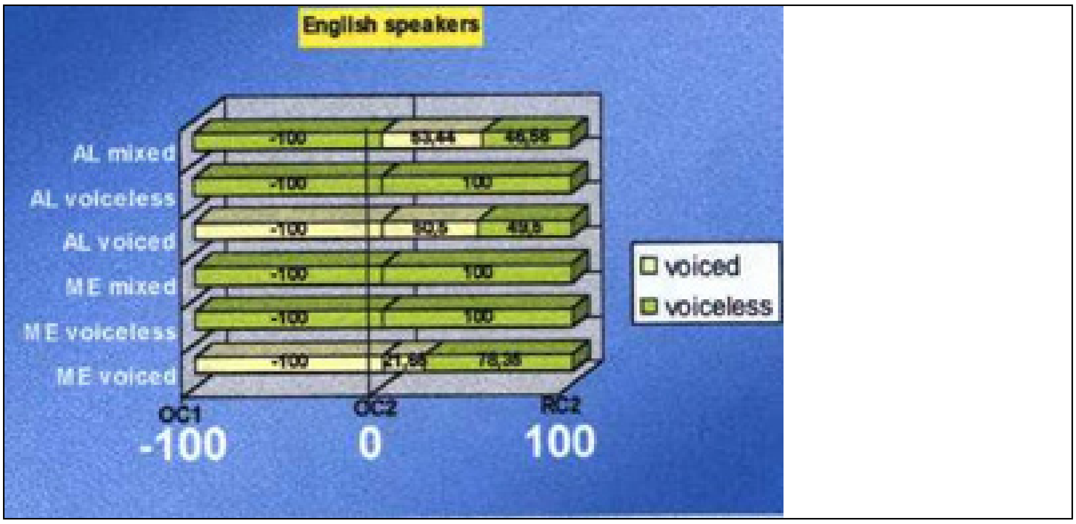

Figure 1.5. Voicing timing patterns found in the control and mixed sequences uttered by the English speakers ME and AL (Cuartero 1998).

Concerning the influence of speaking rate on voicing in English voiceless – voiced stop sequences, Cuartero (1998) found that devoicing of C2 was not dependent on speaking rate. It was also seen that there was variability within and across speakers.

As far as fricative consonants are concerned, it is widely known that speakers of American and British English produce phonologically voiced fricatives as wholly or partially devoiced, so that voicing during the stricture of close approximation that characterizes these consonants does not last as long as the frication noise created at the oral constriction (Docherty 1992). One possible explanation for this tendency is that voicing and frication are difficult to produce simultaneously (Ohala 1983), since they involve conflicting aerodynamic conditions. On the one hand, voiced fricatives require that subglottal pressure be higher than oral pressure in order to maintain vocal fold vibration. On the other hand, high airflow is needed in order to produce the turbulent airflow that characterizes fricative consonants, which requires a high pressure in the oral cavity. Such finely-tuned aerodynamic conditions often result in the loss of voicing in voiced fricatives. In fact, Stevens *et al*. (1992) have shown that voicing does not last the entire duration of the oral constriction of a fricative, if there are no maneuvers that are aimed at prolonging vocal fold vibration. This also explains why voiced fricatives are not frequent in the world's languages, *vis à vi*s voiceless fricatives, and why, in the languages where there are voiced fricatives, they tend to be devoiced.

Balise and Diehl (1994) claim that voiced fricatives are not frequent because they are not perceptually salient, *i.e.*, voicing decreases the high intensity noise that characterizes fricatives. In English, where voiced fricatives are devoiced, speakers use other cues apart from presence or absence of vocal fold vibration in order to cue voicing in fricatives. One of these 'alternative' cues is duration of the constriction (Stevens *et al.* 1992).

Docherty (1992) found three patterns in word-initial fricatives. The first was that the complete duration of the medial phase of the fricative consonant was fully voiced, a pattern which was found in voiced fricatives only. The second pattern was complete voicelessness, which was found mainly in voiceless fricatives, although there were also some phonologically voiced fricatives that showed this pattern. Thirdly, partial devoicing was found in post-vocalic voiceless fricatives, with short voicing continuation, and also in voiced fricatives, where voicing continuation was significantly longer.

In final fricatives three patterns were found, as well. Firstly, full voicing occurred in few cases, and most of these were phonologically voiced fricatives. Secondly, complete devoicing was found in both voiceless and voiced fricatives. Thirdly, the most common pattern in voiced and voiceless fricatives was voicing continuation, with voicing dying out during the medial phase of the fricative. Voicing continuation was significantly longer in voiced fricatives than in voiceless fricatives.

Docherty's (1992) experiment shows that, although voiced obstruents have more voicing than voiceless obstruents, it is not the case that all voiced obstruents are phonetically voiced and all voiceless obstruents are phonetically voiceless, since the former are often completely or partially devoiced, and the latter frequently have voicing continuation into their medial phase. Thus, there is overlap in voicing timing patterns between sounds which are typically labeled as voiced or voiceless (Docherty 1992: 129), and variability within and across speakers seems to be the rule in voicing in obstruents.

Smith (1997) carried out an experiment using acoustic and laryngographic data in order to investigate the devoicing of /z/ as a function of the preceding or following sound, and as a function of prosodic position, that is, whether it was word-final, sentence-final or syllable-final. Electroglottography was used to measure vocal fold vibration in the consonant. In her analysis of voicing, Smith (1997) found great variability between speakers in their propensity to devoice fricatives (see Figure 1.6).

One of the speakers (S1) produced /z/ mostly as a fully voiced consonant, whereas another speaker (S2) produced mostly a fully devoiced fricative. The other two subjects produced the sound with partial devoicing and other patterns, too. Smith (1997) claims that this variability could not be explained on the basis of speaking rate, since the speakers who spoke most rapidly were those who produced partially devoiced fricatives, and the speakers who produced voiced fricatives with either complete voicing or complete devoicing spoke more slowly.

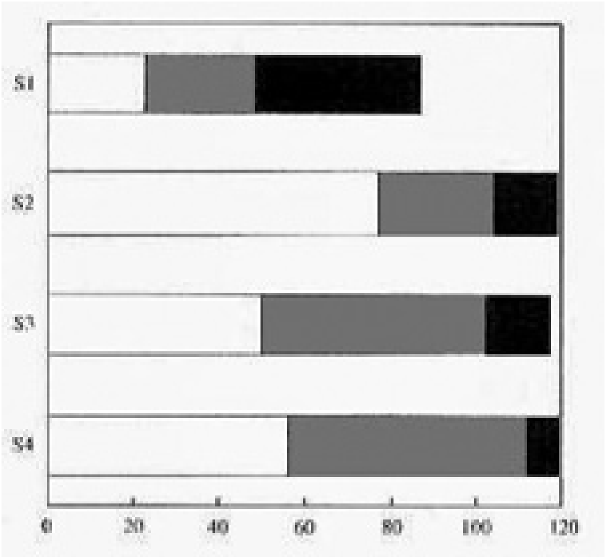

Figure 1.6. Total number of tokens (on the abscissa) that each speaker (on the ordinate) produced as devoiced (white), partially devoiced (gray), or fully voiced (black). Figure from Smith (1997).

Secondly, the results of Smith (1997) also showed that the phonologically voiced fricative /z/ showed a shorter duration than its voiceless counterpart, /s/, in all contexts and for all speakers except for one. The vowel preceding the voiced fricative was also found to be longer than the vowel preceding the phonologically voiceless fricative. Vowel duration was greater for vowels preceding devoiced fricatives than for vowels preceding voiced fricatives. As Smith (1997) suggests, this may be due to the fact that, in the absence of the glottal vibration cue, the larger durational differences aid the perception of the difference between /z/ and /s/.

In addition, Smith (1997) also investigated the effect of segmental context on fricative devoicing. It was found that speakers only produced fully voiced fricatives when the sonorant consonant /l/ followed. Complete devoicing was more frequent when a voiceless stop followed than when the immediately following sound was voiced; the anticipation of the abductory glottal gesture for the following voiceless consonant may explain more extreme devoicing.

The effect of boundaries on devoicing was also measured by Smith (1997). At word boundary, most speakers produced /z/ with partial devoicing, and there were virtually no occurrences of a fully voiced fricative in this context. Thus, devoicing can be attributed to aerodynamic factors.

Interestingly, Docherty (1992) found that in fricative-nasal consonant sequences there was partial devoicing of the nasal stop. In these sequences, there was a lengthening of the interval during which the vocal folds were not vibrating. Besides, the fricative had a shorter duration in this context than in a fricative-vowel context, a phenomenon which results from the compression of the supralaryngeal gesture of the fricative which has been shown to occur in consonant sequences (Haggard 1978). It remains to be seen whether this phenomenon also happens in Catalan sequences.

In fricative-lateral sequences, Docherty (1992) found a longer duration of the noise interval than when fricatives occur in isolation, and he interprets this phenomenon as a delay in voice onset at the onset of the sonorant, although he has no way to draw the boundary between the obstruent and the lateral consonant. This problem may be overcome by using electropalatography, so the experiment that will be described in the forthcoming chapters will shed some light on this phenomenon.

In a second experiment, Docherty explored the effect of the boundary between the two components of a consonant sequence on the timing of voicing. The context analyzed consisted in stop-sonorant sequences and in s-nasal sequences in different boundary conditions. As far as voiceless stop – sonorant sequences are concerned, it was found that the deeper the boundary between the consonants the fewer the cases of voice onset delay. However, the author acknowledges that voice onset delay was consistently not present only when there was a sentence boundary, whereas in the other boundary conditions there was a lot of variability in the occurrence of this phenomenon.

RELEVANCE AND AIMS OF THE STUDY

The study of assimilation has implications for crucial theoretical issues. A fine-grained analysis of assimilation may indicate how languages differ, if they do, in the phonetic realization of alleged phonologically equivalent sequences of sounds. Thus, careful analysis can show how languages differ in the co-ordination of articulatory movements. In the case of voicing assimilation, it is widely known that not all languages coordinate glottal and supraglottal gestures in the same way, and that the same phonological specification may be phonetically implemented differently in different languages.

Secondly, an accurate description of the coordination of oral and glottal gestures may be useful for speech technology. Thus, the results of the present experiment may be used by speech synthesis and speech recognition systems for the creation of natural speech.

Thirdly, cross-linguistic research on voicing assimilation is scant, though it has implications for second language learning. The identification of differences of phonetic implementation across languages in phonologically equivalent sequences of sounds can be used to predict potential errors of transfer from L1 to L2.

Thus, the aim of this dissertation is to use original data from an experiment on the timing of voicing in English and in Catalan to contribute to the following objectives:

a) To provide a quantitative account of the timing of glottal and supraglottal gestures in English and Catalan in consonant sequences across word boundaries;

b) To evaluate current descriptive frameworks;

c) To provide information that can be useful for technological and educational applications.

The aim of the present experiment was to determine the direction and temporal extent of voice assimilation in Catalan and English consonant sequences, and the effect of varying speaking rate and articulatory overlap on assimilation of voicing across word boundaries.

In Catalan, it was expected that voicing assimilation would be complete and regressive. As for speaking rate and articulatory overlap, the prediction was that voicing assimilation would be a categorical process, independent of rate of speech or overlap of supraglottal articulators.

In English, we expected to find partial or complete devoicing of C2 in voiceless – voiced obstruent sequences and partial or complete devoicing in C1 in voiced – voiceless sequences (Jones 1960, Westbury 1979, Docherty 1992, Smith 1997). Furthermore, it was predicted that the assimilation process would be gradient and co-varying with speaking rate and articulatory overlap.

2. Method

2. Method

The present experiment was intended to analyze the process of voicing assimilation in Catalan and English using simultaneous laryngographic, electropalatographic and acoustic data. More specifically, it aimed at observing assimilation when two consonants that have a different specification for voicing occur across word boundaries.

We addressed the following questions: firstly, we intended to observe the direction and the extent of voicing assimilation in Catalan and English. Secondly, the experiment aimed at investigating whether degree of articulatory overlap has an effect on voicing assimilation. Finally, we wanted to know whether speaking rate has any effect on voicing assimilation in both languages.

2.1. Material

The material used in the experiment consisted of meaningful two-word sequences that contained consecutive obstruents or sequences of obstruent and nasal or lateral consonants across word boundaries within a phonological phrase. Two sequence types were studied. On the one hand we analyzed sequences where the first consonant was phonologically voiceless and the second was phonologically voiced (*e.g.* Catalan ‘po/**t#g**/al’, English ho/**t#g**/irl). On the other hand, we studied sequences in which C1 was phonologically voiced and C2 was phonologically voiceless (*e.g.* Catalan ‘ma[**g#t**]ou’, English ‘bi/**g#t**/own’). In the control sequences, the two consonants had the same phonological specification for voicing (*i.e.*, both were voiced or voiceless). Only consonants whose articulatory constriction involved tongue-palate contact – and which were consequently observable in the electropalatographic frames – were used. This means that bilabial stops were not included. Homorganic sequences of obstruents were not analyzed in this experiment because it was not possible to identify the boundary between C1 and C2 in the EPG frames or the acoustic signal, since C1 is obligatorily unreleased when followed by another stop. Figure 2.1 below shows that the same words were used in different combinations in the control and in the test sequences.

The number of syllables and the accentual pattern of the word sequences were kept constant. The vowel context was identical for the control and the test sequences (*e.g.*, ‘sad cap; sad gap; fat cap’) though not always for reverse sequences (*e.g.,* ‘thick lap, full cap’). Each sequence contained two monosyllabic words –except for the adjectives ‘zíngar’ and ‘simple’ in Catalan, which are bisyllabic –with the tonic stress falling on the second word. The material was read twice at three speaking rates –slow, normal and fast- by the Catalan and the English speakers.

The two-word sequences were inserted in the frame sentence *digui* ___ *dos cops* (‘say___twice’) for Catalan, and *say* ___ *again* for English. Both the test sequences and the control sequences are listed in Figure 2.1 below.

The grammatical status and prosodic boundary across the sequences used in the experiment was a factor that could have had an influence on the results. Therefore, the syntactic status of the two-word sequences was kept constant: all of them were noun phrases consisting of a noun preceded by an adjective or a determiner. The natural syntactic order of these two elements in the two languages was respected. Thus, the adjective followed the noun it modified in the Catalan utterances, whereas the modifying adjective preceded the head noun in the English sequences. The same words with the same stress pattern were used in different combinations when possible. The slightly different frequency of occurrence and predictability of the resulting phrase may have affected the results.

		English	*Catalan*
TEST SEQUENCES	/t#g/	fat gap	dret gal
	/d#k/	sad cap	fred car
	/s#d/	this doll	pas dur
	/z#t/	his toll	gas turc
	/l#k/	full cap	sol car
	/k#l/	thick lap	Mac lent
	/_#t/	long toll	fong turc
	/k#n/	thick nut	Mac nou
	/s#n/	this nut	pas nou
	/_#s/	long sip	fong simple
	/s#l/	this lap	pas lent
	/l#s/	full sip	sol simple
CONTROL SEQUENCES	/d#g/	sad gap	fred gal
	/t#k/	fat cap	dret car
	/z#d/	his doll	gas dur
	/s#t/	this toll	pas turc
	/g#n/	big nut	mag nou
	/_#d/	long doll	fong dur
	/z#n/	his nut	gas nou
	/_#z/	long zip	fong zíngar
	/g#l/	big lap	mag lent
	/l#g/	full gap	sol gal
	/z#l/	his lap	gas lent
	/l#z/	full zip	sol zíngar

Figure 2.1. Test and control sequences used in the experiment.

The test sequences were of 4 types:

1) Obstruent sequences where C1 was phonologically voiceless and C2 was phonologically voiced, *e.g.*, ‘fat gap’, ‘this doll’.

2) Obstruent sequences where C1 was phonologically voiced and C2 was phonologically voiceless, *e.g.*, ‘sad gap’, ‘his toll’.

3) Sequences of sonorant plus voiceless obstruent, *e.g.*, ‘full cap’, ‘long sip’.

4) Sequences of voiceless obstruent plus sonorant, *e.g.*, ‘thick lap’, ‘this nut’.

Rather than reverse sequences (*e.g.*, /d#k/ - /k#d/), which would involve variation of voicing and place specification, sequences such as /d#k/ and /t#g/ were analyzed, that is, sequences that involved only variation in voicing.

Different voiceless – voiced and voiced – voiceless sequence types were studied.:

a) Sequences of stop consonants (*e.g.,* /τ#γ/ Catalan ‘dret gal’, English ‘fat gap’, and /d#k/ Catalan ‘fred car’, English ‘sad cap’).

b) Fricative-stop sequences (*e.g*., /σ#δ/ Catalan ‘pas dur’, English ‘this doll’, and /z#t/ Catalan ‘gas turc’, English ‘his toll’).

c) Nasal stops in combination with obstruents (*e.g*., /κ#ν/ (Catalan ‘Mac nou’, English ‘thick nut’ and /σ#ν/ Catalan ‘pas nou’, English ‘this nut’).

d) Lateral consonants in combination with obstruents (*e.g*., /κ#λ/ Catalan ‘Mac lent’, English ‘thick lap’ and /σ#λ/ Catalan ‘pas lent’, English ‘this lap’).

Table I. Sequences studied. The columns identify C2 and the rows identify C1. The symbol '√' means that the sequence was analyzed. 'C' stands for 'control sequence' and 'T' stands for test sequence.

C2 C1	Vls alv stop	Vd alv stop	Vls vel stop	Vd vel stop	Vls fric	Vd fric	Nasal	Liquid /l/
Voiceless alveolar stop			√ C	√ T				
Voiced alveolar stop			√ T	√ C				
Voiceless velar stop							√ T	√ T
Voiced velar stop							√ C	√ C
Voiceless fricative	√ C	√ T					√ T	√ T
Voiced fricative	√ T	√ C					√ C	√ C
Nasal	√ T	√ C			√ T	√ C		
Liquid /l/			√ T	√ C	√ T	√ C		

2.2. Subjects

Four adult subjects participated in the experiment, two native speakers of Catalan and two native speakers of English. All the subjects were linguistically trained, and no subject reported a history of speech or hearing disorders. The Catalan subjects were native speakers of the Eastern variety of Standard Catalan, and they were labeled AN and MJ. They were born and bred in the Barcelona area. They both lived permanently in Barcelona and used Catalan for all their daily linguistic exchange. As for the English subjects, AL spoke the variety of English known as RP and ME spoke General American. It has to be pointed out that the choice of subjects was constrained by the availability of custom-made artificial palates.

2.3. Experimental procedure

Simultaneous acoustic, electropalatographic and electroglottographic data were collected for the subjects reading the speech material. Subjects were asked to sit in front of a microphone. The lights were turned off in order to reduce external noise to the minimum. The electroglottograph (EGG) and the palatograph (EPG) were adjusted to the speakers. Then the subjects were asked to read the test and the control sequences inserted in a frame sentence.

All the utterances were stored in a Pentium III Hewlett Packard Pavilion. Electropalatography was used to trace the movement of the oral articulators, and electroglottography was used to trace the activity of the vocal folds during the production of the consonant sequences. The audio signal of each of the utterances was obtained as well for further acoustic analysis.

The data acquisition and analysis was carried out with the help of the Reading EPG3 software, which provides a simultaneous multi-channel representation of the EPG, the EGG and the acoustic waveform. The sequences were recorded at a sampling rate of 10000 Hz. The duration measurements were double-checked by means of spectrograms obtained with CSL/Multispeech by Kay Elemetrics.

The speech material was read at three different speaking rates. Firstly, the subjects were asked to read the sequences at a normal speed. After that, the subjects were asked to read the sequences more slowly. Finally, the subjects had to read the sequences as fast as they could.

2.3.1. Electropalatography

Electropalatography is a technique that records the timing and location of tongue contacts with the hard palate during running speech. In the Reading EPG3 System, sixty-two miniature silver electrodes that are arranged on the surface of a custom-made artificial palate detect the tongue-palate contact. These palates are made from acrylic, and are molded to fit the subjects' upper palate. The patterns of tongue-palate contact over time are transmitted to the PC computer for storage and display. The display covers one hundred linguo-palatal contact frames per second.

The electrodes are arranged in eight horizontal rows, which correspond to anatomical landmarks, as shown in Figure 2.2.

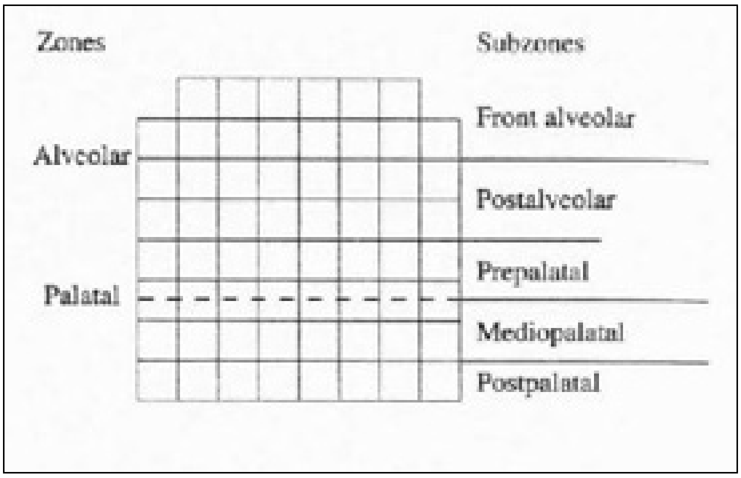

Figure 2.2. Articulatory zones on the electropalate (from Recasens *et al.* 1996)

The frontmost row has six electrodes and the other seven rows have eight electrodes. The surface of the palate is divided into four major zones: alveolar (rows 1-2), post alveolar (rows 3-4), palatal (rows 5-7), and post palatal (row 8). The speech signal was recorded on the left channel while the larynx output was recorded on the right channel. The EPG data is synchronous with data corresponding to acoustical and physiological parameters (EGG). The subjects wore the artificial palates for approximately 10 minutes before the recording.

Concerning this speech analysis technique, it is relevant to mention the study by Fougeron *et al.* (2000), which compares the Kay Elemetrics and the Reading EPG systems. The study compares how variations in tongue-palate contact are reflected by the two systems and evaluates the contribution of the additional 34 electrodes of the Kay electropalatograph in terms of description of the articulation of speech sounds.

The study mentions three main differences between the Kay and the Reading EPG systems. Firstly, the plate containing the electrodes is smaller in the Reading system; it covers only the hard palate and stops at the gingival border, as opposed to the Kay system, which goes further back in the oral cavity towards the soft palate. Secondly, the Reading artificial palate is thicker than the Kay palate -1.5 millimeters for Reading and 1 millimeter for Kay –so that the sensory feedback seems to be better with the Kay palate. However, the study mentions the fact that adjustment to the Reading palate improves greatly after a training period. Thirdly, a very important difference between both systems is the number and placement of the electrodes on the pseudo-palates. The Reading system has 62 electrodes arranged in eight horizontal rows, whereas Kay has 96 electrodes arranged in arches around a mid-sagittal line. The Kay system has a better coverage of the dental area and there are more electrodes in the velar region. Additionally, there are also more electrodes in the palatal region.

The corpus they studied included the French consonants /n, k, l, s/ and the results showed that Kay provided additional information for all the consonants except for /l/. Additional information in the Kay system appears mainly in the dental and velar regions, due to the difference in electrode coverage between the two palates. However, their study concludes that, although the Kay system has a better definition of the articulation of segments, the same articulatory effects can be shown by both palates. The conclusion, then, is that, although less accurate than the Kay system, the Reading EPG3 system is accurate enough to describe the contacts of the tongue with the hard palate in the production of the consonant sequences that are the object of this study.

2.3.2. Electroglottography

Electroglottography was used in this experiment because it accurately detects vibration of the vocal folds. Besides, it is a non-invasive technique, so that it allows the recording of natural samples of the subject's behavior.

The Electroglottograph used in this experiment was an FJ Electronics, type EG – 830. Electroglottography is a technique that is used to measure the variations in the electrical impedance between the vocal folds. The impedance is sensed by means of a weak high frequency alternating current of 300,000 Hz, which is sent through the larynx from two silver electrodes that are placed on each cricoid cartilage. When the vocal folds vibrate, the area of contact between the right and the left vocal fold changes. In this way, the electrical impedance between the two electrodes changes, too. The result is that the vocal fold movements and vibrations modulate the alternating current between the two electrodes.

The EGG signal is rather unstable and is subject to the physiological conditions of the informants. Firstly, movements of the head may cause variations in the thickness of the fat tissue between the silver electrodes and the larynx. To solve this problem, the informants were asked not to move their head during the recording sessions. In addition, frequent pauses were made in order to avoid tiredness. Secondly, it is very difficult to keep the contact resistance between the glottograph electrodes and the skin at a constant value because ultra slow variations in skin humidity cause variations in contact resistance. To overcome this problem, conducting EGG jelly between the electrodes and the skin was applied to the subjects. Figure 2.3 shows a sample printout displaying the simultaneous acoustic, EPG and EGG signals in the stop sequence /t#g/.

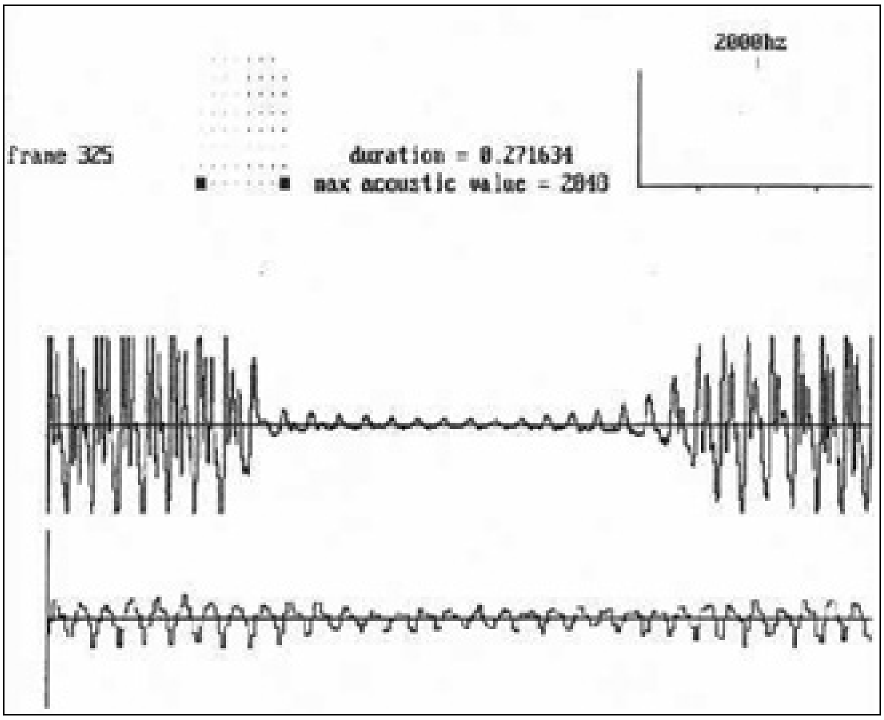

Figure 2.3. Sample printout showing the stop sequence /τ#γ/ in the Catalan utterance *pot gal,* displaying the EPG signal (top left), the acoustic waveform (middle), and the EGG trace (bottom).

2.4. Analysis procedure

The experiment had both a between-subject and a within subject design. We wanted to test whether there are cross-linguistic differences in the implementation of voicing in consonant sequences. Additionally, we intended to measure the effect of speaking rate on voicing in the languages studied, namely English and Catalan, because if voicing is a phonetic process in English as opposed to a phonological process in Catalan, different effects of speech rate on assimilation are to be expected (Solé 1995a).Voicing was considered to be actual vocal fold vibration as reflected in the EGG trace.

The EPG contacts were used to determine the onset and offset of the articulatory constriction for C1 and C2 in the consonant sequences that were the object of this study. A frame-by-frame representation of the articulation of the consonant sequences in each utterance was obtained with the Reading EPG3 software (see Figure 2.4 below) . Thus, the tongue-palate contacts at intervals of 10 milliseconds could be observed.

As illustrated in Figure 2.4, the following points were identified in the frame-by-frame printout of the consonant sequences following Gibbon, Hardcastle, Nicolaidis (1993):

1) Approach to C1 (AC1): the onset of movement of the tongue towards the maximal constriction of C1, as shown on the EPG signal.
2) Onset of C1 (OC1): the point in time at which the total or maximal constriction for the first consonant in the sequence began.
3) Release of C1 (RC1): the point at which the articulators started moving apart for the release of C1 and the release burst on the acoustic signal if there was any.
4) Approach to C2 (AC2): the onset of movement of the tongue toward the maximal constriction of C2.

5) Onset of C2 (OC2): the point in time at which the total or maximal constriction for the second consonant in the sequence began.
6) Release of C2 (RC2): the point at which the articulators started moving apart for the release of the second consonant in the sequence and the release burst on the acoustic signal if there was any.
7) Point of Maximal Constriction (PMC): used to determine the onset of a consonant when no complete constriction was present.

These annotation points were then used to determine which portions of the consonant sequence showed vocal fold vibration. The interval in the EGG waveform representing glottal excitation was measured relative to the annotation points and the percentage of voicing for each consonant was then calculated. As illustrated in Figure 2.4, time 0 in the consonant sequence was defined as onset of the constriction for C2 where a switch in the state of the glottis (voiced to voiceless or voiceless to voiced) would be expected if there was perfect synchronization between the oral and glottal articulations. Thus, regressive voicing assimilation was expressed in terms of negative percentages, whereas progressive assimilation of voicing was expressed in terms of positive percentages. If the percentage of voicing was zero, it meant that there was no assimilation; in other words, each of the consonants in the sequence maintained its original phonological specification for voicing (*e.g.*, /dk/ realized as [dk]). A positive value indicated progressive assimilation of vocal fold vibration (*e.g.*, /dk/ realized as [dgk]) or lack of vocal fold vibration (*e.g.*, /tg/ realized as /tg g/). A negative value indicated regressive assimilation of vocal fold vibration (*e.g.*, /tg/ realized as [tdg]) or absence of vocal fold vibration (*e.g.*, /dk/ realized as [dtk]). Additionally, voicing continuation into the stop sequence closure was measured, and the percentage of voicing in C1 and C2 was calculated.

One-way ANOVAs were performed with sequence type (voiced, voiceless, and mixed) as the independent variable and percentage of voicing in the consonant type as the dependent variable. Post-hoc tests (Scheffé) allowed us to locate significant differences. The statistical analysis was carried out with Microsoft Excel 2000 and SPSS 10.0.

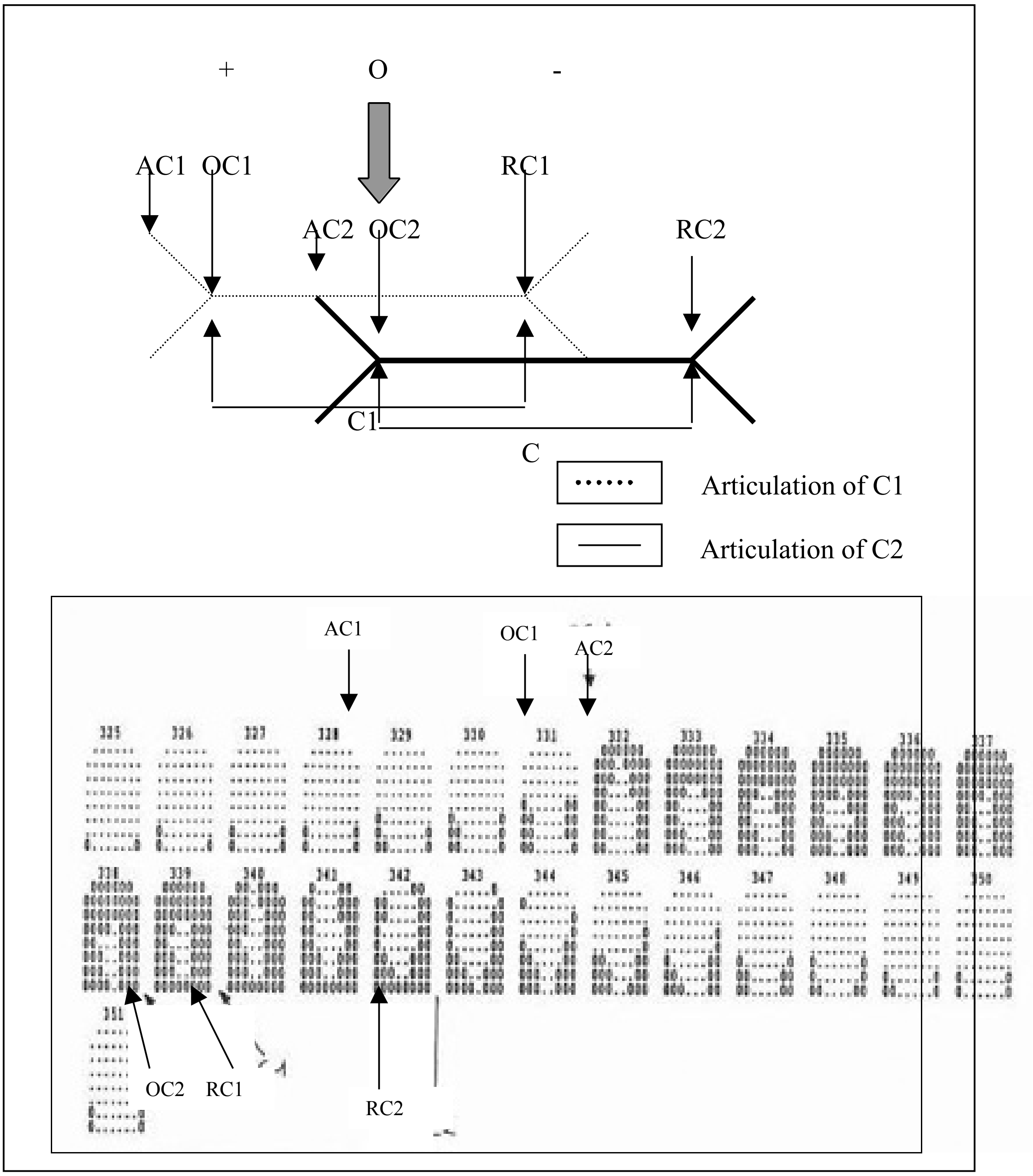

Figure 2.4. Schematic representation of the labeling and measuring criteria in the consonant sequences (top) and EPG printout (bottom) of the consonant sequence /t#g/ from the Catalan sequence *dret gal* with the annotation points marked.

Figure 2.5 below displays the consonant sequence /d#k/ in ‘fred car’ from the Catalan speaker AN with the measurement points marking segmentation of the signal. The Figure displays the EPG signal (top), acoustic waveform (middle) and EGG signal (bottom). The measurement points represent AC1 (approach to C1), OC1 (onset of C1), RC1 (release of C1), AC2 (approach to C2), OC2 (onset of C2) and RC2 (release of C2). Sample data from the rest of the test sequences can be found in Appendix C (page 215).

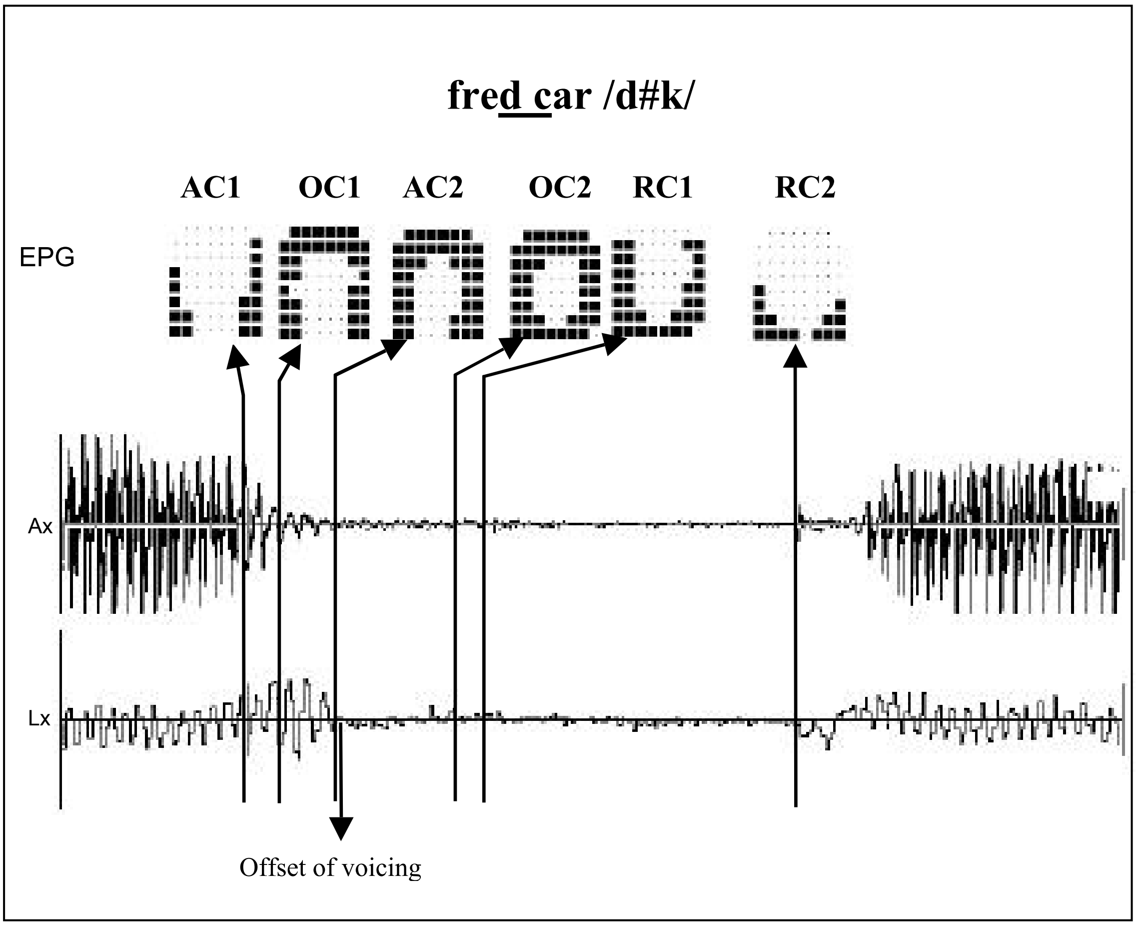

Figure 2.5. Sample sequence from the Catalan sequence /d#k/ in ‘fred car’ (uttered by subject AN) with the measurement points marking segmentation of the signal. The figures display the EPG signal (top), acoustic waveform (middle) and EGG signal (bottom) corresponding to the test sequences. The measurement points represent AC1 (approach to C1), OC1 (onset of C1), RC1 (release of C1), AC2 (approach to C2), OC2 (onset of C2) and RC2 (release of C2).

In addition voicing in obstruents before and after a pause was measured in order to compare them to consonant sequences.

Additionally, the index of overlap of the consonant constrictions and the consonant gestures was computed. Index of overlap of the closure (OIOC) was calculated from time measurements using the annotation points as follows (Gibbon *et al.1993*):

$$\mathrm{OIOC} = (\mathrm{OC2}-\mathrm{RC1} * 100) / \mathrm{OC1}-\mathrm{RC1}$$

If OI= 100, then the closure of C2 completely overlapped the closure of C1. If OI= 0, then there was no overlap but perfect sequencing of the two constrictions.

The index of overlap of the articulatory movements of C1 and C2 (OIACO) measured the degree of overlap of the articulatory movements for the two consonants in the sequence. The index was calculated from time measurements using the annotation points as follows:

$$\mathrm{OIACO} = (\mathrm{AC2}-\mathrm{RC1} *100) / \mathrm{AC1}-\mathrm{RC1}$$

The EPG signal and the waveform were used to measure the duration of the frame sentence, which served as an index of speaking rate. The criteria for the measurements of speaking rate were the following:

- In Catalan utterances, the duration of *digui* and *dos cops* was measured.
- In the English utterances, the duration of the words *say* and *again* was measured.

Pearson’s correlation tests were carried out in order to see whether there was a significant effect of speaking rate and articulatory overlap on percentage of voicing in the test sequences.

SUMMARY

Summing up, the following measurements were made for each utterance:

i) Duration of C1 (OC1-RC1) and movements for C1 (AC1-RC1).

ii) Duration of C2 (OC2-RC2) and movements for C2 (AC2-RC2).

iii) Duration of the voiced and voiceless portions of the consonants.

iv) Percentage of voicing in C1 and C2.

v) Index of overlap of the closure (OIOC).

vi) Index of overlap of articulatory movements toward C1 and C2 (OIACO).

vii) Duration of the frame sentence.