GLOSSOLOGIA 11-12 (2000) 167-199

ACOUSTIC STRUCTURE OF THE GREEK STOP CONSONANTS*

ANTONIS BOTINIS, MARIOS FOURAKIS & IRINI PRINOU

The Greek stop consonants [p/b], [t/d] and [k/g] were analysed in the vocalic context [o]. The durations of the consonantal occlusion, release and aspiration (if any) were measured, and spectral measurements were made of the consonantal release and the first two formants of the adjacent vowels. The results showed that the voicing of the occlusion as well as the durations of the occlusion and aspiration provide acoustic evidence for voice and place of articulation oppositions. The occlusion duration of the voiced stops was found to be significantly longer than that of their cognate voiceless stops and the distribution of the aspiration measurements showed structural differences for both the presence vs. absence of voicing, and the different places of articulation, i.e. velar>labial>alveolar.

1. Introduction

1.1. Purpose of study and objectives

The present study is an acoustic analysis of the Greek stop consonants in standard Athenian Greek. Two main questions are addressed: (1) what is the acoustic structure of the voiced vs. voiceless oppositions and (2) what is the acoustic structure of the labial vs. alveolar vs. velar places of articulation? More specific questions are addressed with reference to voice and place of articulation acoustic correlates which involve consonantal as well as adjacent vowel acoustic characteristics. The consonantal characteristics refer to tempo-

---

* The present study is based on work in progress on the phonetic structure of Greek at the University of Athens. The authors benefited from discussions with Robert Bannert and colleagues at the Phonetics Department of Umea University, Sweden, where the main results were presented at a seminar by Antonis Botinis in September 1996. Jan van Doorn, University of Sydney, Australia, and Gunnar Fant, Royal Institute of Technology, Sweden, made content as well as editorial comments on the final version of the paper for which the authors express their sincere appreciation.

ral patterns of the occlusion, release and aspiration as well as spectral patterns of the release whereas the vocalic characteristics refer to formant patterns and formant transitions from the nucleus into the offset of the pre-consonantal vowel and from the onset into the nucleus of the post-consonantal vowel.

1.2. Related studies of the Greek stop consonants

Most phonological studies in Modern Greek (henceforth simply Greek, except in contrasting reference to Classical Greek) include a description of the stop consonants and are usually based on a structural analysis of their underlying phonological representations (e.g. Newton 1961, 1972; Koutsoudas 1962; Householder 1964; Warburton 1970 and Pagoni 1993). There is still, to our knowledge, hardly any published acoustic data on Greek stop consonants. The question addressed by most studies, and a major issue of Greek phonology, has been the phonemic status of the voiced stop consonants. Newton (1961, 1972), Warburton (1970) and Pagoni (1993) argue for the non-phonemic status of the voiced [b], [d] and [g] stops and suggest an underlying representation of a two-phoneme sequence, /mp/, /nt/ and /nk/ respectively, whereas Koutsoudas (1962) and Householder (1964) attribute a phonemic status to the voiced stops and propose a paradigmatic opposition between /p/, /t/, /k/ and /b/, /d/, /g/. A nasal + a voiceless stop may not surface as a voiced stop before obstruents, e.g. ['pempti] 'Thursday', ['lampsi] 'gleam' whereas loan words may show stylistic variation and surface as a nasal + a voiceless stop. However, regardless of the issue of derivation (from underlying to surface representation), there are minimal pairs, contrasting in voicing, such as labial ['kopos] 'fatigue' vs. ['kobos] 'knob', alveolar ['patos] 'bottom' vs. ['pados] 'anyhow' and velar ['pakos] 'pack' vs. ['pagos] 'bench'. The voiced stop may be in free variation with a nasal + voiced stop realisation, i.e. ['kobos] or ['kombos], in accordance with the personal index, speaking style, dialect, etc.

1.3. The stop consonant system of Greek

Modern Greek has a 2-way stop system, voiced vs. voiceless, in comparison to the 3-way system of Classical Greek, voiced vs. voiceless unaspirated vs. voiceless aspirated. Other languages may also have a 2-way system but different distinctions, e.g. voiceless unaspirated vs. voiceless aspirated or even a 4-way system, e.g. voiceless unaspirated vs. voiceless aspirated vs. voiced unaspirated vs. voiced aspirated (see e.g. Lisker and Abramson 1964; Maddieson 1984). The Modern Greek voiceless stops are derived from the Classical Greek voiceless stops whereas the voiced stops have been mostly derived from Classical Greek nasal+stop combinations, e.g. {en#poros}=> ['eboros] 'merchant', {en#'tonos}=>['edonos] 'intense', {en#krisis}=>[egrisi] 'approval'. Classical Greek aspirated stops, on the other hand, have turned into voiceless fricatives, e.g. {'p^{h}ilos}=>['filos] 'friend', {'t^{h}ema}=>['θema] 'subject', {'k^{h}aos} =>['xaos] 'chaos' whereas voiced stops have turned into voiced fricatives in Modern Greek, e.g. {'bios}=>['vios], {'drama}=>['ðrama], 'drama', {'logos}=> ['loɣos], 'speech'.

The Modern Greek writing system has adopted the majority of the Classical Greek conventions. Thus, structural phonetic changes by the 4th century AD may not be reflected in Modern Greek orthography. Diphthongs, which have turned into monophthongs, may be digraphs, short vs. long quantity distinctions, which have merged into short quantities, may be different letters, and syllabification, which has turned into an open syllable structure, may allow double consonant letters in accordance with Classical Greek orthography. On the other hand, aspiration and accentual markers (i.e. grave and circumflex) may not be used in standard Modern Greek orthography any more except for an acute mark, usually on the vowel of the stressed syllable. The voiceless stops have not undergone any major phonetic changes from classical times and are represented by the same letters (π, τ, ϰ) and neither have the "double" classical consonants, i.e. {ξ} and {ψ}, which correspond to two consonantal sounds, i.e. /ks/ and /ps/ respectively. The voiced stops are represented by two letters as a rule, i.e. {μπ}, {ντ} and {γϰ} or {γγ}, for labials, alveolars and velars respectively (and to a limited extend {νδ} for alveolars), which reflects, to some degree, the phonetic and etymological diachrony of the Modern Greek vocabulary. On the other hand, the letters {β}, {γ} and {δ} represent the Modern Greek fricatives, i.e. /v/, /γ/, and /ð/ respectively, which derive from the respective homorganic Classical Greek voiced stops. Thus, the alphabetic letters {β}, {γ} and {δ} are pronounced ['vita], ['ɣama] and ['ðelta] respectively in Modern Greek and not ['bɛta], ['gama] and ['delta], as they have been adopted by the majority of the western European languages with reference to the Classical Greek pronunciation.

Greek consonants may form clusters at the syllabic onset but not at the syllabic coda positions where the open syllable structure of the language syllabifies intervocalic consonants to the right as a rule. Thus, there are hardly any words of Greek origin with stop endings, although loan words and foreign names, expressive and onomatopoetic words and abbreviations may end in consonants other than the regular [s] or [n] endings. Stops may be branched with all other classes of consonants on the right, including stops, but not fricatives. Onset stop clusters may undergo dissimilation, which turns the first stop of the cluster into a homorganic fricative, e.g. ['ktima]=>['xtima] 'estate', although Classical Greek clustering may resist dissimilation and may even produce minimal distinctions in the Modern Greek vocabulary, e.g. [ptiˈno] ~ [ftiˈno] 'bird ~ cheap'. In a two-node consonantal clustering, stops may be branched with sibilants (s/z) on the left, in addition to stops and fricatives (cf. above) whereas, in a three-node clustering, stops may be branched by glides or nasals on the right if the first constituent is [s], e.g. [stroˈfi] 'turn', [ˈsplina] 'spleen', [ˈsknipa] 'gnat'.

Both voiced and voiceless stops may be involved in palatalisation processes in unstressed prosodic contexts where the underlying post-consonantal /i/ has a wide range of surface realisations, in accordance with the preceding consonant (e.g. [kuˈpi] ~ [kuˈpça] 'oar ~ oars', [kuˈbi] ~ [kuˈbja] 'button ~ buttons', [kuˈti] ~ [kuˈtça] 'box ~ boxes', [ˈɣadi] ~ [ˈɣadja] 'glove ~ gloves', [kuˈci] ~ [kuˈca] 'bean ~ beans', [puˈɟi] ~ [puˈɟa] 'purse ~ purses'). Palatalisation may not apply, even in optimal segmental and prosodic contexts (e.g. [ˈipios] 'mild', [ˈetios] 'responsible', [ciˈanio] 'cyanide'), due to morphological, etymological or sociolinguistic conditions.

2. Background and theoretical framework

2.1. Segmental structure of the stop consonants

Stop consonant production is characterised by three main successive phases: (1) a total closure of the vocal tract at some place of articulation (e.g. labial, alveolar, velar), which causes a temporal cessation of airflow through the vocal tract (hence the term "stop" or the Greek term "kleista", i.e. closed); (2) a sudden opening of the vocal tract by means of a rapid movement of the articulators apart, which causes a momentary airflow burst at the consonantal release of the corresponding place of articulation; (3) a delay of vowel onset, during which the vocal cords are into a quasi-open position and air flow through the glottis causes aspiration noise. The successive phases of stop production are referred to as "occlusion", "release" and "aspiration" in the present study whereas the term "burst" is reserved for both release and aspiration as one structural unit. Traditional articulatory descriptions, based on the movements and states of the articulators associated with the stop consonant production, refer to the vowel-to-consonant and consonant-to-vowel articulatory movements: "The formation of a stop normally has three phases: the *onset* of closure, when an articulator is approaching the other; the *closure*, when the articulators are held together, obstructing the airflow; and the *offset* of the closure, when the articulators are moving apart again." (Henton and Ladefoged

1992, p. 66). A common term, which denotes the explosive nature of the stops is "plosives" whereas the term "obstruents" includes stops and fricatives as opposed to "sonorants" (vowels and vowel-like sounds such as "semi-vowels").

Standard acoustic analysis of the stop consonants (e.g. Fant 1960, 1968, 1973) measures five successive segments: (1) occlusion (voiced or silent), (2) transient (i.e. release burst), (3) fricative segment, (4) aspirative segment and (5) the initial part of a following voiced sound (if it is coarticulated with the stop). Thus, Fant (1973) distinguishes between "fricative segment" and "aspirative segment" in principle although the acoustic realisation may be overlapping and frication and aspiration may thus merge into an aspiration-like segment: "The aspirative segment can in part cooccur with the fricative segment but takes over as the degree of articulatory opening proceeds." (Fant 1973, p. 113). On the other hand, release, frication and aspiration are usually treated and referred to as a single acoustic unit, namely "burst".

Stop consonants are the most widely studied speech sounds as they form complicated phonetic structures which are highly coarticulated with the adjacent sounds and have thus raised competitive theories about the processes of speech production and perception, as well as the acoustic correlates of the stop consonant distinctions. For example, some 16 acoustic correlates have been reported for the voiced vs. voiceless stop distinctions in the literature and about 7 are often referred to for the place of articulation distinctions (see e.g. Edwards 1981; Lisker 1986). The onset of voicing has been widely discussed for the voiced vs. voiceless oppositions whereas the spectral pattern in the vicinity of the consonantal release as well as the formant transitions (i.e. the formant trajections into the adjacent vowel) have drawn the attention of most investigators for the place of articulation oppositions.

2.2. The voicing distinction

The voiced vs. voiceless opposition has been studied for a wide range of languages (e.g. Lisker and Abramson 1964; Fant 1973; Edwards 1981; Abramson and Lisker 1995; Larranaga, Lleo and Prinz 1995) the results of which show considerable variability in different contexts as well as among languages. Thus although the occlusion may be voiced or voiceless in accordance with the voiced vs. voiceless opposition respectively for some languages and contexts, other languages and contexts may not utilise the voiced vs. voiceless realisation for this opposition. Instead of the voiced vs. voiceless realisation of the occlusion, the Voice Onset Time (VOT), i.e. the onset of voicing in relation to the consonantal release has been suggested as a single acoustic dimension for the voicing distinction across languages (Lisker and Abramson 1964; Abramson and Lisker 1995).

VOT is assigned a zero-time value at the reference point of the consonantal release, negative values for voicing before the release and positive values for voicing after the release. Negative VOT is typical for voiced stops with quasi-periodic voicing during the occlusion and is said to have voicing "lead" whereas positive VOT is typical for voiceless stops with variable degrees of aspiration and is said to have voicing "lag". In classificatory terms, voicing lead, short voicing lag and long voicing lag are assumed to correspond to voiced, voiceless unaspirated and voiceless aspirated stops respectively. Thus, VOT is assumed to classify related acoustic correlates in a straightforward way in such languages as e.g. English, German or Swedish where voicing and aspiration may be combined in different ways in different phonetic contexts.

Apart from VOT, which covers the distribution of release and aspiration, occlusion duration, total duration, post-consonantal voice fundamental frequency (F_0) and post-consonantal vowel duration have been reported as acoustic correlates of stop voicing distinctions (see e.g. Edwards 1981). The occlusion duration may be longer for voiced stops than their cognate voiceless but this difference has not drawn much attention and is not an established acoustic fact in the literature on different languages. The total duration of the voiceless consonants, especially in languages with aspiration, may be longer than their voiced cognate. The F_0, which is the physical correlate of the voice pitch, is higher for a vowel succeeding a voiceless stop than a voiced stop and the duration is longer for a vowel preceding a voiced stop than a voiceless stop, other segmental and prosodic effects being equal. Voiced and voiceless stops have also been referred to as "fortis" and "lenis" (or "tense" and "lax") respectively, where the voiceless stops are assumed to be related to greater articulatory force than the respective voiced ones (see Jakobson and Halle 1962). The association of voicing with articulatory force has been seriously questioned (see Fant 1960; Lisker and Abramson 1964) and the respective terms are hardly in use in current phonetic theory.

2.3. Place of articulation

Place of articulation oppositions have also been studied thoroughly for different languages where most research has focused on the spectral shape of the release and the formant transitions (see e.g. Fant 1973; Stevens and Blumstein 1978; Blumstein and Stevens 1979, 1980; Edwards 1981; Kewley-Port 1982, 1983; Kewley-Port, Pisoni and Studdert-Kennedy 1983; Lahiri, Gewirth and Blumstein 1984; Repp and Lin 1989; Sussman 1991; Sussman, McCaffrey and Matthews 1991; Sussman, Hoemeke and Ahmed 1993; Cassidy and Harrington

1995). The spectral shape at the vicinity of the consonantal release has been reported with three basic patterns which correspond to three places of articulation: Diffuse-falling (or diffuse-flat), diffuse-rising and compact for labial, alveolar and velar stops respectively. Labial and alveolar stops are rather flat with no major spectral peak and are thus referred to as "diffuse" whereas velar stops have one spectral peak at the midfrequency region which dominates the entire spectrum and are thus referred to as "compact" (hence the distinctive feature terms "compact" vs. "diffuse"). On the other hand, the falling (or level) spectral pattern of the labial stops with higher energy distribution at lower frequencies is referred to as "grave" in contrast to the rising pattern of the alveolar stops with higher energy distribution at higher frequencies which is referred to as "acute" (hence the distinctive feature terms "acute" vs. "grave").

The formant transitions are highly context-sensitive and may have a large variability in different vocalic contexts. F1 has minimal correlation with place of articulation realisation (but it may indicate the onset of VOT), F3 also has a marginal one, and higher formants, such as F4, are not considered to have any systematic variation. Thus, in practice, F2 transitions reflect in the first place the notion covered by "formant transitions". A formant transition is traditionally defined by its onset frequency, its offset or steady-state frequency and its duration (see e.g. Delattre, Liberman and Cooper 1955; Kewlly-Port 1982). Early work at Haskins Laboratories emphasised the significance of formant transitions for place of articulation distinctions (e.g. Cooper, Delattre, Liberman, Borst, and Gerstman 1952; Liberman, Dellatre and Cooper 1952; Liberman, Dellatre, Cooper, and Gerstman, 1954; Liberman, Cooper, Shankweiler and Studdert-Kennedy 1967). Although formant transitions in CV-syllables were reported to be different for each place of articulation depending on the immediate following vowel, the formants, especially F2, started from or pointed at a specific point of the frequency axis, i.e. the "locus" for each place of articulation. Thus, labial, alveolar and velar stops, the last ones in a nonback vowel context, are assumed to have a locus at the vicinity of 720 Hz, 1800 Hz and 3000 Hz respectively.

The significance of formant transitions in the classification of stop consonants is also acknowledged in early work by Stevens and Blumstein (e.g. Stevens and Blumstein 1978; Blumstein and Stevens 1979, 1980). However, formant transitions are presumed by some investigators to contribute secondarily and in combination with the release spectrum which is assumed to be the primary acoustic cue (cf. above). It is further assumed that release spectrum and formant transitions are interdependent acoustic properties and are thus hypothesised as an integrated cue for place of articulation oppositions (see also Lieberman and Blumstein 1988). Subsequent work on formant transitions (e.g. Fant 1973 on Swedish and Kewley-Port 1982 on American English), reported much larger variability and overlapping patterns between the different places of articulation than the distribution of locus frequencies would predict and alternative acoustic correlates were proposed. The locus concept has however been redefined and recent work reports on formant transitions as reliable acoustic correlates which are referred to as "locus equations", i.e. regression lines derived by the computation of F2 onset, F2 target (vocalic steady-state) and F3 onset (Sussman 1991; Sussman *et al.* 1991; Sussman *et al.* 1993). The significance of formant transitions to place of articulation oppositions was also argued in Cassidy and Harrington (1995) who reported good overall classification scores at vowel onset frequencies. The contribution of formant information to place classification is greater for F2 than for F3 than for F1 and, furthermore, the release spectrum seems to provide more information than the formant transitions do. However, both the release spectrum and the formant transitions provide sufficient information and thus, according to Cassidy and Harrington, support a model for "sufficient discriminability".

Apart from release spectral shapes and formant transitions, the duration of VOT, the duration of the occlusion, and the strength of the release amplitude have been reported, among others, as acoustic correlates of stop voicing distinctions (see e.g. Edwards 1981). Thus, although VOT may be a primary acoustic correlate for the voicing distinctions, it has been found to increase in absolute value going from the labial to the alveolar to the velar places of articulation (Lisker and Abramson 1964; Edwards 1981; Kewley-Port 1982). VOT has been reported as the primary acoustic correlate for place of articulation oppositions, clearly overriding the classificatory effects of formant movements: "VOT was found longer as place of articulation moves from the front of the vocal tract to the back, i.e. labial < alveolar < velar, and this acoustic correlation appears to be invariant across different vowels" (Kewley-Port 1982). The duration of the occlusion has also been reported to vary in accordance with the place of articulation, i.e. labials > alveolars > velars, which is in the opposite direction to, and in complementary distribution with, VOT (Zue 1976; Umeda 1977; Edwards 1981). The release amplitude has been reported strongest for the alveolar and weakest for the labial stops and this separation may be more categorical if labial stops are included into one category with lower release amplitude and non-labial stops are included into another category with higher release amplitude (Zue 1976; Edwards 1981).

2.4. Production aspects

Stops are the class of consonants that exist in all languages (Jakobson and

Halle 1956; Maddieson 1984; Henton, Ladefoged and Maddieson 1992). They involve two basic characteristics which distinguish them from all other sounds: a total closure of the vocal tract and a succeeding abrupt release. Both closure and release result in radical changes in the acoustic signal, i.e. silence and burst, which are extremely salient for the auditory system and create the most favourable markers for speech perception.

Stop production requires precise interarticulator programming which involves complex physiological coordination of the larynx, soft palate, tongue, lips and jaw. For voiced stop production, laryngeal adductor muscles (mainly the lateral croicoarytenoid and the interarytenoids) close the glottis (i.e. the opening between the vocal cords). The transglottal air pressure differential created by the respiratory system (mainly the intercostal muscles and the diaphragm) forces the vocal cords apart and myoelastic and aerodynamic forces (enforced by the "Bernoulli" effect) bring the vocal cords together, to repeat the cycle and produce thus a periodic glottal buzz (Van der Berg 1958; Sonesson 1970). For voiceless stop production, on the other hand, the vocal cords are wide open by the action of the abductor laryngeal muscles (i.e. the posterior croicoarytenoids, which are the only abductor laryngeal muscles). There is thus no vibration of the vocal cords, which corresponds to a silent interval during occlusion, and the first audible sound is created at the point of articulation upon the release of the occlusion. During aspiration, the vocal cords are quasi-open so that aspiration (h-like sound) is generated at the glottis and the air flow has a free passage throughout the vocal tract. The soft palate is typically raised and blocks the nasal cavity passage preventing air flow through it, whereas the tongue, lips and jaw regulate the closure of the vocal tract in various ways and result in occlusions at different places.

The stop release is a characteristic of both voiced and voiceless stops, although for the voiced ones the release amplitude is weaker as a result of the drop in air pressure conditioned by the laryngeal activity of the vocal cords. Voicing and aspiration, on the other hand, are associated with voiced and voiceless stops respectively whereas early and late maximal glottal opening during closure may be correlated with non-aspirated and aspirated voiceless stops respectively (Lofqvist 1980).

2.5. Perception aspects

Research on stop consonant perception, initiated at Haskins Laboratories in the 1950s with synthetic speech (e.g. Cooper *et al.* 1952; Liberman, Delattre, Cooper and Gerstman 1954; Delattre, Liberman and Cooper 1955), has created a long-lasting debate on the nature of speech perception. The main approach of this early research was that phonetic segments (stops) are not processed perceptually as beads on a string but contextual information conditions their classification. The complexity of the speech signal and the special nature of speech have been referred to as the "speech code", and the variability of the acoustic signal and its relation to production mechanisms is referred to as speech "encoding" whereas the perception of the speech signal with reference to production mechanisms is referred to as speech "decoding". Thus, neither the release spectrum nor the onset frequencies or the direction of the formant transitions are assumed to provide invariant acoustic cues but, rather, place of articulation classification is primarily determined by the adjacent vowel context. The large variability of the speech signal and the lack of one-to-one correspondence between the acoustic signal and the phonetic percept led to the development of the "motor theory" of speech perception, i.e. a theory of speech perception in which the acoustic signal is interpreted at a higher level in terms of abstract features with reference to the speech production system (e.g. Liberman, Cooper, Shankweiler and Studdert-Kennedy 1967; Studdert-Kennedy, Liberman, Cooper and Harris 1970; Dorman, Studdert-Kennedy and Raphael 1977; Liberman and Mattingly 1985).

Another argument of early perceptual research with reference to the special nature of speech has been that there are distinct perceptual modes for different phonetic categories. Stop consonants and vowels are assumed to be processed in a "categorical" and a "continuous" mode respectively. Thus experiments with synthetic stimuli have reported an abrupt identification change across stop category boundaries whereas identification change across vowel category boundaries is continuous. On the other hand, discrimination scores for pairs of stimuli from different stop categories were high whereas discrimination scores for pairs of stimuli belonging to the same stop category were close to chance. On the contrary, vowel stimuli belonging to the same vowel category may be discriminated to a much higher degree than chance. Vowels are thus processed like all other non-speech sounds, i.e. in a continuous (i.e. non-categorical) mode. Further evidence for the special nature of speech has been provided by "dichotic" listening experiments. In these experiments a non-speech sound is played to one ear and a speech sound is played to the other ear. When the speech sound is processed by the right ear and the non-speech sound by the left ear, which are connected to the left and right hemisphere of the brain respectively, the intelligibility of speech is high. However, intelligibility drops substantially when the speech sound is processed by the left ear and the non-speech sound by the right ear. Thus, it seems that the left hemisphere is adapted to speech processing more than the right hemisphere, and this distinguishes speech and its properties from all other sounds.

Although the lack of phonetic invariance for distinctive speech sounds as well as the categorical nature of speech are well established research paradigms, alternative approaches on these issues have been developed. Thus, invariant acoustic patterns have been reported for stop consonant distinctions of place of articulation in the vicinity of the consonantal release by Stevens and his associates (e.g. Stevens and Blumstein 1978; Blumstein and Stevens 1979, 1980, see *Place of Articulation* above). This approach is mainly based on the predictions of Fant's theory of speech production (Fant 1960), according to which different vocal tract shapes will produce distinct acoustic patterns.

2.6. Cross-linguistic classifications

Stops show great cross-linguistic variability with reference to both voicing and place of articulation. Some 7 phonation types have been reported (Ladefoged 1988; Henton *et al.* 1992) from which the most common ones are voiced, voiceless and aspirated. Keating (1984) refers to these three types as "major phonetic categories". Other phonation types involve breathy voice (murmur), slack voice, creaky voice (laryngealised), and stiff voice. The unmarked phonation type is the voiceless one. Ruhlen (1975) reports 706 languages out of which 166 have solely voiceless stops in comparison to only 4 with solely voiced stops whereas the majority of languages, i.e. 536 languages, have both voiced and voiceless types of stops. There are some 16 separate places of articulation reported (Henton *et al.* 1992), and for the 317 languages included in the UCLA Phonological Segment Inventory Database the alveolar/dental is the most common place of articulation followed by the velar place but only with a marginal difference. Other places of articulation in a descending order involve bilabial, palatal (or palato-alveolar), uvular, retroflex, and labial-velar whereas the remaining of the places of articulation are used by less than 6% of the languages (Maddieson 1984).

3. Experimental design and methodology

3.1. Speech material

The speech material consists of meaningful disyllabic test words in the carrier sentence [i 'leksi ___ 'ine elini'ci] 'the word ___ is Greek'. The test words have a (C)VCVC segmental structure, where the vocalic segments are the mid, back vowel [o] and the intervocalic consonant under investigation is one of [p/b, t/d, k/g] which is syllabified to the right and thus forms the onset of the second syllable. The test words are: [ˈopos] ‘(such) as’, [ˈkobos] ‘knob’, [ˈsotos] ‘Sotos’ (proper name), [ˈodos] ‘really’, [ˈkokos] ‘speck’, and [ˈogos] ‘mass’. The speech material was written down with each sentence on a separate piece of paper and was read seven times, in a random order each time, with normal tempo and no particular focus or prosodic break. The first and last repetitions were rejected for the elimination of any prosodic interference and the remaining five repetitions were acoustically analysed, yielding thus a total of 120 utterances (4 speakers x 5 repetitions x 3 places of articulation x 2 voicing conditions).

3.2. Speakers

The speakers were four male students who were between 20 and 25 years old. All four speakers have been brought up in Athens and speak standard Athenian Greek. They have not lived abroad and speak one or two foreign languages (mostly English).

3.3. Experimental equipment

The speech material was recorded at the Athens University Phonetics Laboratory, under quiet conditions, with a TC-WR590 Sony cassette deck and a dynamic unidirectional Shure Prologue microphone with 40 Hz to 13 kHz frequency response. The material was analysed in the Kay Elemetrics CSL environment (hardware and software) at Athens University with the default settings of CSL at 10 kHz sampling rate and 20 ms frame length with 16 bit resolution (for the burst analysis the tokens were resampled at 20 kHz and the frame length was set to 10 ms).

3.4. Acoustic analysis and measurements

The acoustic analysis was carried out with reference to the occlusion, release and aspiration of the stop consonants as well as the transitions of the first two formants of the adjacent vowels in accordance with the following 8 measurement points indicated in Figure 1:

1. Nucleus formant frequencies of the preceding vowel, i.e. F1 and F2 frequencies at five glottal pulses from the offset of the vowel.
2. Formant offset of the preceding vowel, i.e. F1 and F2 frequencies at vowel offset as evidenced by the offset of its full-range formant tracing.
3. Occlusion duration, i.e. from the offset of the preceding vowel up to

the release onset.

4. Release spectrum, i.e. release spectral shape and formant energy distribution within a 10 ms window from its onset.
5. Release duration, i.e. from the onset up to the release offset.
6. Aspiration duration, i.e. from the release offset up to the onset of the vowel as evidenced by the onset of its full-range formant tracing.
7. Formant onset of the following vowel, i.e. F1 and F2 frequencies at vowel onset.
8. Nucleus formant frequencies of the following vowel, i.e. F1 and F2 frequencies at five glottal pulses from the onset of the vowel.

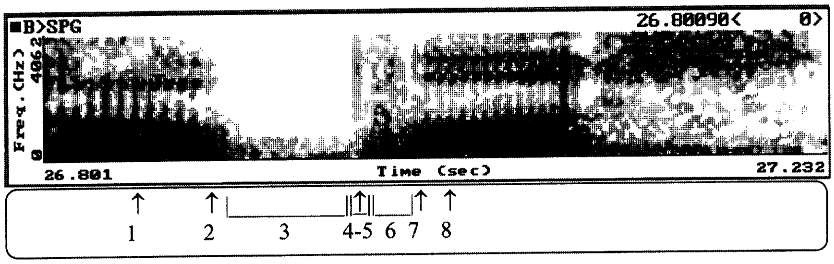

FIGURE 1: Measurement points with reference to spectral and temporal realisations of [p] in the lexical context [ˈopos] (see text).

The acoustic analysis procedures were carried out in the CSL environment, University of Athens, for display, measurements, and playback of the signal. Duration measurements were made by numerical cursor readouts at measurement points with reference to the waveform aligned with the time-spectral wide spectrogram display. Spectral measurements of the release were made by numerical cursor readouts of Linear Predictive Coding (LPC) at release onset with reference to the spectrogram display. The formant measurements of the vowels were made at measurement points with reference to the spectrogram display as follows: (I) by eye inspection monitored by cursor numerical display on the screen, (II) by an LPC-derived formant history superimposed on wideband spectrogram and monitored by cursor numerical display, and (III) by numerical cursor readouts at peak frequencies of LPC spectral analysis.

3.5. Statistical analysis and processing

Standard statistical procedures for the analysis of speech were followed and the acoustic results were subjected to the Analysis of Variance with the Statview statistics software package at the Ohio State University Speech and Hearing Department.

4. Results

4.1. Durational characteristics

Figure 2 shows an example and Table I shows the durations of voiced vs. voiceless and labial vs. alveolar vs. velar place of articulation oppositions. From the occlusion, release and aspiration measurements, the VOT and the total duration of the stop consonants are derived. Means are shown across speakers (beginning of table) and for each speaker separately.

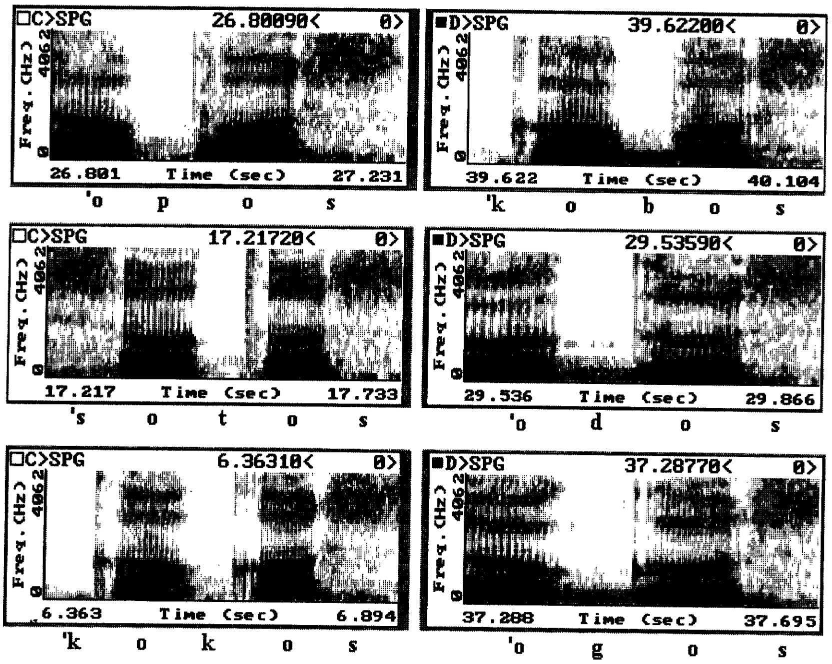

FIGURE 2: One speaker's wide band sample examples of ['opos], ['kobos], ['sotos], ['odos], ['kokos] and ['ogos] (see text).

The total duration grand means of the stop consonants do not show any systematic differences either for the voiced vs. voiceless or the labial vs. alveolar vs. velar oppositions; the duration of all consonants are between 89.9 ms and 94.5 ms except for the voiceless alveolar [t] which is 84.6 ms. However, the occlusion duration and the aspiration duration showed noticeable differences for the voicing opposition but minor differences for the place of articulation oppositions. The voiceless and the voiced stops varied from 62.5 ms to 65.0 ms and 78.3 ms to 81.5 ms respectively. A two factor analysis of variance

	OCCLUSION		RELEASE		ASPIRATION		BURST	VOT	TOTAL
	XG	S	XG	S	XG	S	XG	XG	XG
p	63.0	7.1	8.3	0.5	18.6	4.3	26.9	+26.9	89.9
t	62.5	11.6	6.9	0.5	15.2	4.1	22.1	+22.1	84.6
k	65.0	4.5	9.5	1.7	19.6	5.6	29.1	+29.1	94.1
b	79.8	7.0	11.4	3.4	-		11.4	-79.8	91.2
d	81.5	8.3	11.2	1.7	-		11.2	-81.5	92.7
g	78.3	8.7	16.2	1.5	-		16.2	-78.3	94.5
Speaker 1									
p	59.6	9.6	8.6	0.8	15.0	6.8	23.6	+23.6	83.2
t	54.6	7.7	7.2	0.4	13.4	3.6	20.6	+20.6	75.2
k	64.4	5.7	8.4	1.1	11.6	3.9	20.0	+20.0	84.4
b	81.2	8.2	9.4	1.5	-		9.4	-90.6	90.6
d	76.8	5.5	9.4	1.5	-		9.4	-86.2	86.2
g	72.8	6.1	15.4	2.8	-		15.4	-88.2	88.2
Speaker 2									
p	59.4	5.9	8.0	4.2	24.6	8.3	32.6	+32.6	92.0
t	71.8	11.3	7.6	1.6	14.6	4.8	22.2	+22.2	94.0
k	69.8	14.8	10.8	4.0	20.0	4.3	30.8	+30.8	100.6
b	73.0	11.3	9.2	1.6	-		9.2	-73.0	82.2
d	87.2	21.3	13.0	2.1	-		13.0	-87.2	100.2
g	82.8	12.7	18.0	11.4	-		18.0	-82.8	100.8
Speaker 3									
p	59.4	3.9	8.0	4.2	24.6	8.3	32.6	+32.6	92.0
t	50.4	6.9	6.2	1.0	21.2	1.0	27.4	+27.4	77.8
k	59.0	7.8	7.6	0.8	24.4	2.8	32.0	+32.0	91.0
b	89.0	11.7	16.6	2.4	-		16.6	-89.0	105.6
d	89.8	17.3	12.4	1.6	-		12.4	-89.8	102.2
g	88.4	7.9	17.0	4.4	-		17.0	-88.4	105.4
Speaker 4									
p	73.8	5.8	9.0	1.0	19.2	1.9	28.2	+28.2	102.0
t	73.2	5.9	6.8	0.4	11.8	3.5	18.6	+18.6	91.8
k	67.0	8.6	11.2	2.1	22.4	5.8	33.6	+33.6	100.6
b	76.0	12.7	10.6	2.1	-		10.6	-76.0	86.6
d	72.2	3.7	10.0	1.2	-		10.0	-72.2	82.2
g	69.4	6.2	15.4	2.8	-		15.4	-69.4	84.8

Table I. Grand means (XG), means (X) and standard deviations (S) of duration patterns (in ms) of 4 speakers x 5 repetitions for voicing and place of articulation oppositions.

showed a significant effect of voicing with voiced stops having longer occlusion durations than the voiceless ones (F(1,114) = 57.689, p <.0001) but no significant effect of place of articulation (F(2,114) = .024, p >.05. There was also no significant interaction between voicing and place of articulation. The occlusion of the voiced stops was fully voiced as a rule whereas that of the voiceless stops showed large variability from partly voiced to voiceless. The release durations varied from 6.9 ms to 9.5 ms and from 11.2 ms to 16.2 ms for the voiceless and voiced stops respectively. The aspiration durations showed remarkable differences between the voiced and voiceless stops but minor differences among the places of articulation. The aspiration duration varied from 15.2 ms to 19.6 ms for the voiceless stops whereas the voiced stops had hardly any aspiration at all. A one-factor analysis of variance for the aspiration duration of the voiceless stops showed no significant effect of place of articulation (F(2,57) = 2.9, p >.05). The calculation of the VOT showed positive values, from 22.1 ms to 29.1 ms, and negative values, from 78.3 ms to 81.5 ms, for the voiceless and voiced stops respectively. The combined release plus aspiration interval, i.e. the burst, ranged from 22.1 to 29.1 ms for the voiceless stops whereas, for the voiced stops, no aspiration interval was measurable and thus the release and burst values were the same, from 11.2 to 16.2 ms. A two way ANOVA revealed a significant effect of voicing, that is this interval was longer for the voiceless stops than for the voiced (F(1.114) = 192, p <.01). There was also a significant effect of place, with velar stops having longest interval, followed by the alveolar and the labial stops (F(2.114) = 13.3, p <.01). There was no significant interaction.

4.2. Energy distribution of the release

Figures 3a and 3b show average spectra for the voiced stop releases computed for each speaker from the five repetitions analysed and, similarly, figures 3c and 3d show average spectra for the voiceless stops.

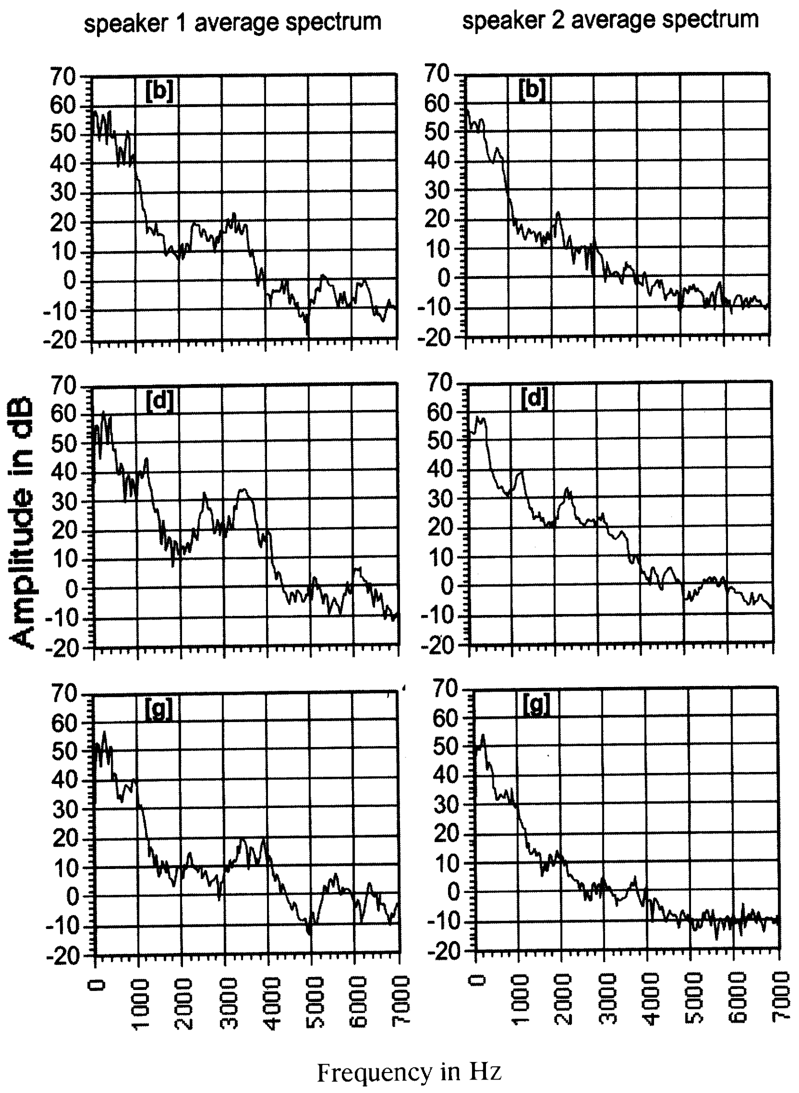

FIGURE 3a: Average spectra for the voiced stop releases [b], [d] and [g] of speaker 1 (on the left) and speaker 2 (on the right).

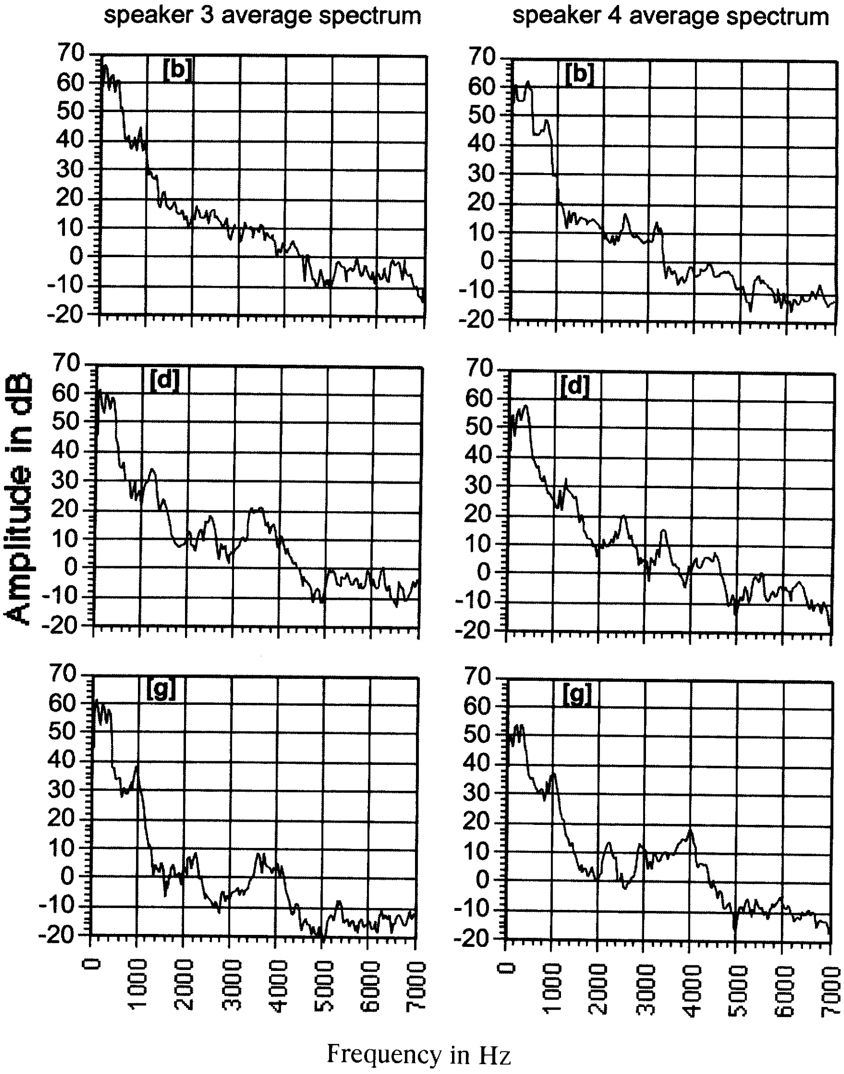

FIGURE 3b: Average spectra for the voiced stop releases [b], [d] and [g] of speaker 3 (on the left) and speaker 4 (on the right).

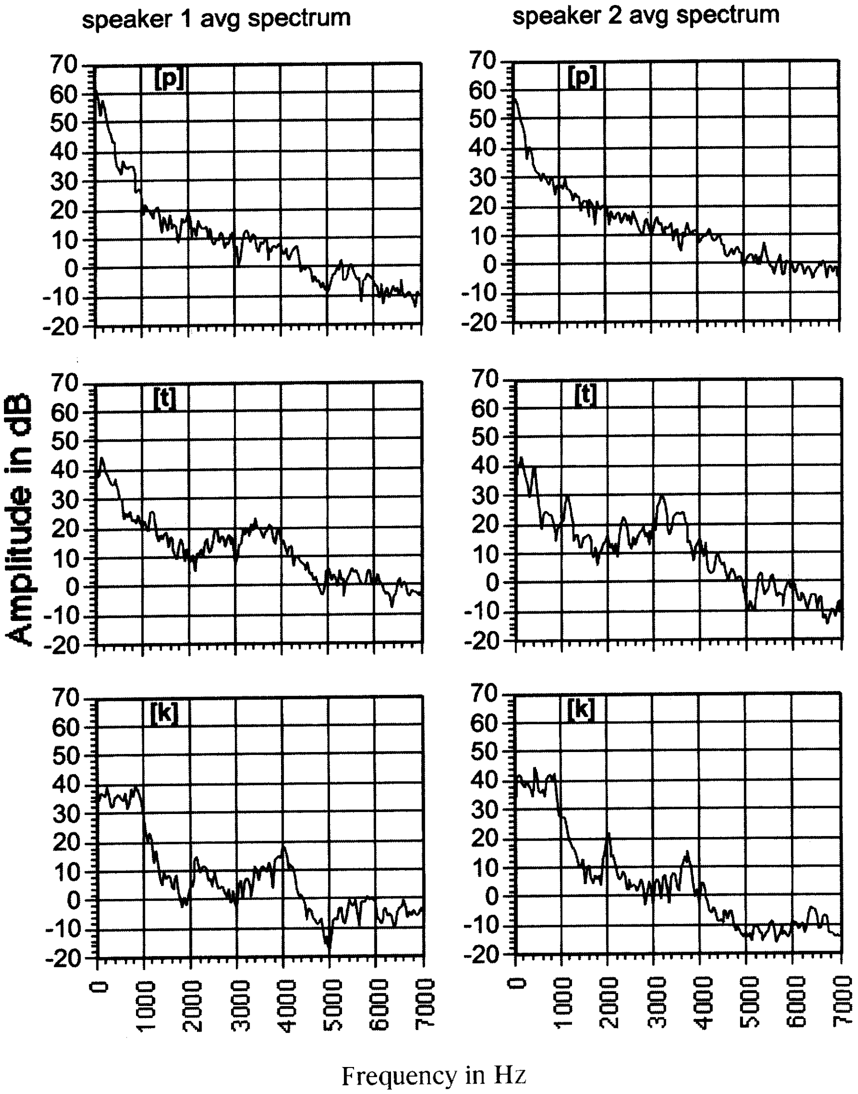

FIGURE 3c: Average spectra for the voiceless stop releases [p], [t] and [k] of speaker 1 (on the left) and speaker 2 (on the right).

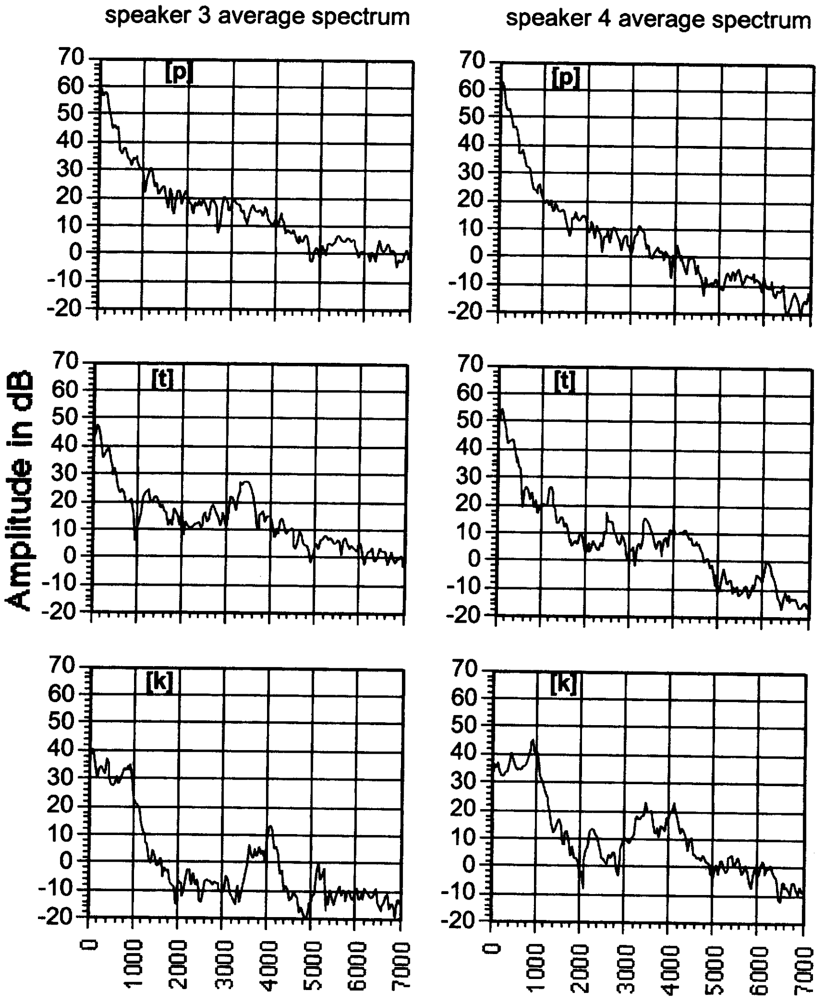

FIGURE 3d: Average spectra for the voiceless stop releases [p], [t] and [k] of speaker 3 (on the left) and speaker 4 (on the right).

All spectra, regardless of voicing or place of articulation show a peak in frequencies below 1000 Hz. This is due to the fact that even though the analysis window for the releases was set at 10 ms, it was in many cases long enough to include some of the voicing of the following vowel. In fact, in the voiced stop spectra, individual harmonics of the glottal source can be clearly seen as in all three voiced stops for speaker 1. Other than this low frequency energy, the different stops exhibit the following patterns:

A. Both voiced and voiceless bilabial stops show a steady decrease in energy going to the higher frequencies, the kind of energy distribution usually called diffuse.

B. The voiceless alveolar stops show a peak in the frequency range between 3 and 4 kHz.

C. The voiced alveolar stops show a peak in the frequency range between 2.5 and 4 kHz.

D. The voiceless velar stops show one peak around 2 kHz and one around 4 kHz.

E. The voiced velar stops show a peak at 1 kHz for 3 of the speakers and another at 4 kHz for all speakers.

In general the spectra for the alveolar and velar stops are not as distinct as the ones reported in the literature for e.g. the American English stops. A more extensive study using more speakers and more vowel contexts might help provide more distinct spectral shapes for each stop.

4.3. Formant transitions

Figure 4 and Table II show the pre- and post-consonantal F1 and F2 distribution for voiced vs. voiceless and labial vs. alveolar vs. velar place of articulation oppositions. The points at which measurements were taken are: A. vowel nucleus; B. vowel offset; C. vowel onset; D. vowel nucleus (see *Acoustic analysis and measurements*).

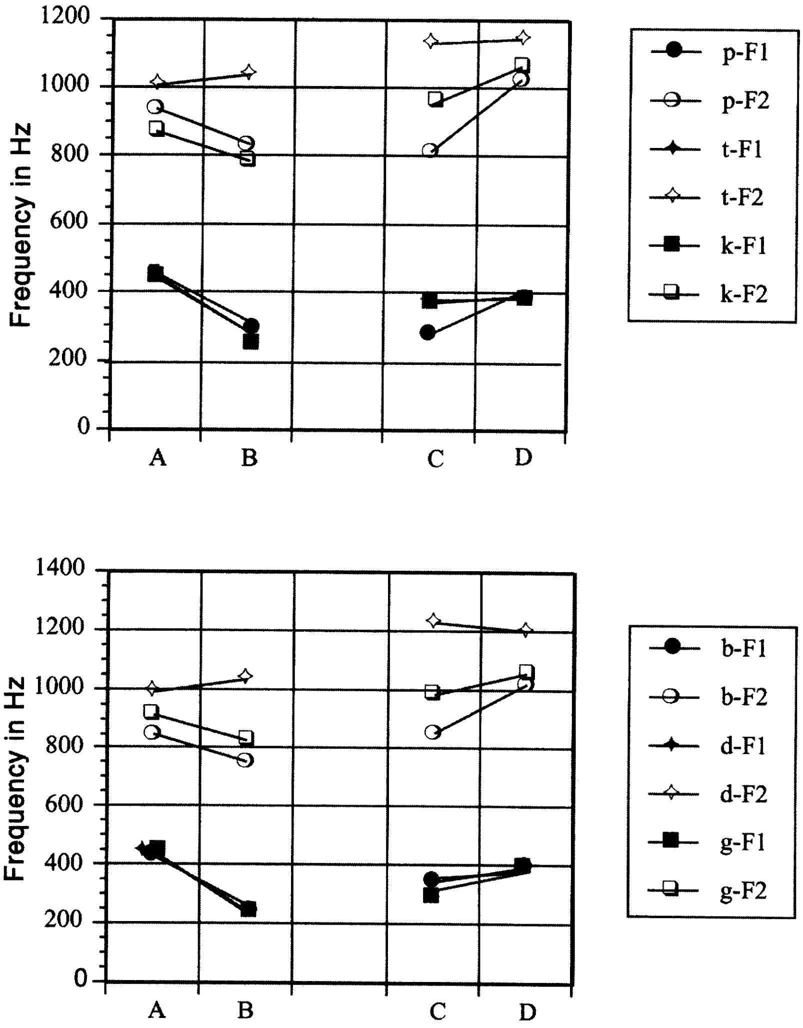

FIGURE 4: Adjacent vowel F1 and F2 frequencies and transitions for [b/d/g] (lower panel) and [p/t/k] (upper panel).

	V1 NUCLEUS				V1 OFFSET				V2 ONSET				V2 NUCLEUS			
	F1		F2		F1		F2		F1		F2		F1		F2	
	XG	S	XG	S	XG	S	XG	S	XG	S	XG	S	XG	X	XG	S
p	453	27	937	84	284	69	841	41	275	71	815	24	404	26	1026	55
t	454	23	1009	61	315	40	1039	75	376	5	1131	38	383	22	1141	30
k	443	26	870	43	282	48	785	78	369	17	953	76	390	8	1059	80
b	433	36	850	53	266	65	751	13	343	27	852	35	396	6	1018	23
d	441	31	992	64	239	31	1030	49	359	24	1225	23	381	18	1197	57
g	443	23	911	86	234	21	825	46	317	26	983	35	378	16	1054	64
Speaker 1																
p	472	15	853	26	383	31	805	43	246	27	818	27	424	9	957	15
t	483	27	950	17	297	99	955	43	374	18	1142	11	383	25	1122	21
k	472	9	822	13	345	0.8	774	14	381	13	908	20	386	15	988	51
b	459	15	787	9	343	8	735	39	378	22	851	10	400	31	1002	32
d	469	18	912	30	263	35	960	29	394	10	1206	22	393	7	1191	31
g	476	45	798	18	255	31	816	32	350	10	997	7	393	7	1046	45
Speaker 2																
p	435	7	944	51	271	43	794	68	201	78	847	10	390	9	1027	34
t	435	7	1033	48	352	40	1019	23	379	11	1076	22	397	17	1119	35
k	421	15	846	29	232	30	715	48	384	12	907	26	400	14	1036	24
b	407	15	825	41	231	52	752	34	347	42	856	54	397	21	1012	49
d	421	9	987	41	263	35	960	29	394	10	1206	22	383	14	1202	52
g	428	8	929	33	213	12	774	45	306	10	939	7	390	9	1026	45
Speaker 3																
p	424	9	1050	25	221	18	887	38	283	27	789	69	374	11	1092	39
t	435	7	1085	37	345	0.5	1137	41	370	9	1137	5	351	9	1185	9
k	421	9	894	14	260	4	755	7	367	15	1068	14	379	12	1175	12
b	397	10	891	9	293	0.4	752	0.4	337	1	894	8	386	9	1053	40
d	407	15	1071	20	269	8	1071	7	340	8	1226	12	355	9	1268	15
g	424	15	1008	61	262	18	823	6	324	3	975	45	365	14	1147	45
Speaker 4																
p	479	7	901	25	262	45	837	47	369	31	804	23	428	19	1029	38
t	463	7	969	32	268	88	1047	55	381	8	1167	39	400	7	1138	30
k	459	26	918	14	293	49	898	8	346	25	931	14	394	7	1037	45
b	469	22	897	19	196	26	767	47	310	20	807	11	400	7	1005	7
d	466	16	996	30	214	15	1060	18	357	6	1257	37	393	8	1128	12
g	445	7	908	12	238	8	887	1	288	1	1022	22	362	24	998	15

Table II. Grand means (XG), means (X) and standard deviations (S) of F1 and F2 (in Hz) of 4 speakers x 5 repetitions for voice and place of articulation oppositions.

The pre-consonantal (V1) and post-consonantal (V2) vowel formant structure is quite similar, with a tendency for the pre-consonantal formant structure of the vowels towards more centralised values. The pre-consonantal F1 grand mean values range from 433 Hz for [b] to 454 Hz for [t] and the post-consonantal ones range from 378 Hz for [g] to 404 Hz for [p] whereas the pre-consonantal F2 grand mean values range from 850 Hz for [b] to 1009 Hz for [t] and the post-consonantal ones range from 1018 Hz for [b] to 1197 Hz for [d].

Figure 4 shows the formant trajectories across speakers for the pre-consonantal and post-consonantal vowels as interpolations between the points measured. The formant transitions show contextual as well as place of articulation structural covariations. The pre-consonantal F1 transition has a falling structure whereas the post-consonantal F1 transition has a rising structure, regardless of the voice vs. voiceless and the place of articulation oppositions. The largest pre-consonantal F1 transition is 209 Hz for [g], from 443 Hz to 234 Hz, whereas the smallest one is 139 Hz for [t], from 454 Hz to 315 Hz. The largest post-consonantal F1 transition is 129 Hz for [p], from 275 Hz to 404 Hz, whereas the smallest one is 7 Hz for [t], from 376 Hz to 383 Hz.

The pre-consonantal F2 has a rising structure for the alveolar place of articulation, from 1009 Hz to 1039 Hz for [t] and from 992 Hz to 1030 Hz for [d] whereas the labial and velar places of articulation have a falling F2 structure, the largest one from 850 Hz to 751 Hz for [b] and the smallest one from 870 Hz to 785 Hz for [k]. The post-consonantal F2 has a rather F1 mirror image structure. The voiceless alveolar [t] has a level F2 transition, from 1131 Hz to 1141 Hz, whereas the voiced alveolar [d] has a rather falling F2 transition, from 1225 Hz to 1197 Hz. On the other hand, the labial and velar stops have a rising F2 transition, from 815 Hz to 1026 Hz, the largest one, for the labial [p] and from 983 Hz to 1054 Hz the smallest one for the velar [g].

5. Discussion

The stop consonant oppositions analysed in the present study refer to consonantal as well as contextual acoustic correlates. The consonantal acoustic correlates include the occlusion, release, aspiration and VOT and the contextual correlates include the formant frequency and formant transitions of the adjacent vowels. In summary, the acoustic correlates of the voicing oppositions in the Greek stop consonants refer to: (1) voicing of the occlusion, (2) duration of the occlusion and (3) duration of burst (including aspiration). The acoustic correlates of the place of articulation oppositions on the other hand refer to: (1) duration of the burst, (2) spectral shape of the release and (3) formant transitions. Neither the total duration of the stops nor the formant frequencies of the adjacent vowel nucleus indicate any significant difference either for the voicing (i.e. voiced vs. voiceless) or the place of articulation oppositions (i.e. labial vs. alveolar vs. velar).

The voicing distinction is mainly realised by the presence versus absence of voicing as well as temporal coordination. The presence or absence of voicing is a rather dominant characteristic in Greek, as the voiced stops are fully voiced and the voiceless stops may vary from partly voiced to voiceless. The temporal realisation is a clear cut acoustic correlate as the occlusion of the voiced stops is significantly longer than their voiceless cognates; on the other hand, the voiceless stops have a noticeable vowel onset delay, the correlates of which include an expansion of the consonantal burst, as evidenced by aspiration, i.e. a VOT lag. Voiced vs. unvoiced occlusion differences are quite common among different languages for the voiced vs. voiceless stop realisation respectively (e.g. French, Hungarian, Dutch), where one finds a voicing lead which corresponds to negative VOT values. Other languages (e.g. English, German, Swedish) may not utilise the voicing realisation for the voice distinction in certain contexts, such as in word initial position. Instead, a voicing lag which corresponds to positive VOT values may be used for the voice distinctions among languages (see Lisker and Abramson 1964; Fant 1973; Davis 1994; Scully and Mair 1995).

The duration of the occlusion was within a 62-65 ms grand mean range, with a 50-73 ms mean range for the voiceless stops, and within a 78-81 ms grand mean range, with a 69-89 ms mean range, for the voiced stops. Thus, the Greek data has shown a significant temporal correlate of the occlusion with an average order of 15 ms for voice distinctions. A much smaller difference between the voiced and voiceless stops has been reported for English, 96 ms and 89 ms for mean occlusion duration for voiced and voiceless stops respectively, but no conclusion has been reached as other studies have pointed to the opposite direction or found no temporal correlate of the voice distinctions (see Edwards 1981). The duration of the occlusion did not show any significant correlation with different places of articulation, although analyses of other languages, mostly English, report an ascending order from front to back, i.e. labials > alveolars > velars (cf. Edwards 1981). Typical values of the occlusion among unrelated languages analysed in the same experimental framework have been reported in the proximity of 100 ms with a 75-125 ms range (Lisker and Abramson 1964).

Aspiration is distributed regularly within a 15-19 ms grand mean range, with a 11-24 ms mean range for the voiceless stops whereas the voiced stops are virtually unaspirated. The stop release as a separate acoustic feature has a structure similar to aspiration and thus burst duration, which corresponds to positive VOT values, and has the following distribution: k > p > t. This order is a general tendency and refers to the grand mean of four speakers and not each speaker individually (see Table I). The VOT has been widely attested in different languages to vary in accordance with the places of articulation, i.e. k > t > p, although alternative orders such as k > p > t have been reported (e.g. Lisker and Abramson 1964; Edwards 1981; Kewley-Port 1982). Most studies on this issue distinguish three hierarchical duration classes: a velar one, an alveolar one and a labial one. Thus the present study supports a hierarchical distribution in theory, although the actual order may vary. Obviously, these results will have to be corroborated with additional Greek material. Typical positive VOT values (which parallel to aspiration) among unrelated languages analysed in the same experimental framework have shown a median of 10 ms (with 0-25 ms range) and 75 ms (with 60-100 ms range) for the voiced (unaspirated) and voiceless (aspirated) stop consonants respectively (Lisker and Abramson 1964). According to this classical taxonomy voiced stops may be associated with considerable aspiration, albeit to a limited degree in comparison to aspiration associated with voiceless stops and, even more, aspiration found in languages where it may have a distinctive function.

The Greek voiceless stops may, in accordance with the classical taxonomy, be characterised unaspirated, as is widely reported (see e.g. Joseph and Philippaki-Warburton 1987). In acoustic terms, however, aspiration is a phonetic reality which, although it is not used for aspiration distinctions, may be used for voice and place of articulation distinctions in combination with other acoustic correlates of the respective categories. The Greek voiceless realisation is similar, at least at some contexts such as word initial, to the English voiced ones. Thus the English, German or Swedish /d/ e.g. may be pronounced as the Greek /t/ (i.e. [t]) and lead to confusion and misunderstanding for minimal pairs in Greek (e.g. /dora/ ~ /tora/ 'Dora' ~ 'now'), having /dora/ uttered as /tora/ by an English, German or Swedish speaker respectively (pointed out by Robert Bannert, personal communication). On the other hand, aspiration is in complementary distribution with voicing in English as a rule (i.e. aspiration is not assumed to be phonemic in English) as in word medial position and/or intervocally the voiced stops may be typically voiced. The general categorisation of both voiced and voiceless Greek stops as unaspirated not only misses a precise description but it also deprives phonetic theory from its acoustic correlates and perceptual cues for phonetic distinctions. Furthermore, the non-utilisation of aspiration for speech synthesis and speech recognition systems may prove a serious drawback for synthetic speech naturalness and speech recognition effectiveness.

As outlined in the "Acoustic analysis and measurements" section, the stop consonants and their segments in this study have been measured (from the spectrograph) with reference to the last and first full-range formant tracings and not the last and first glottal pulse of the preceding and succeeding vowels respectively. This type of analysis is comparable to the classical type of analysis which has the glottal pulses as reference points for measurements. A comparison between the two types of analysis has not resulted in major differences: "for the voicing onset definition, the mean VOT for the voiced stops was found to be 15.0 ms (S = 9.3 ms) and 66.6 ms (S = 16.1 ms) for voiceless stops. For the vowel onset definition, the mean VOT for the voiced stops was 16.7 ms (S = 9.2 ms) and 68.6 ms (S = 16.2 ms) for the voiceless stops." (Edwards 1981, p. 540). This author concludes that the "vowel onset", i.e. the type of analysis of the present study, is closer to Lisker and Abramson (1964) study which is the standard VOT measurements and analysis reference: "The "vowel onset" measurement is perhaps more closely associated with Lisker and Abramson's (1964, p. 389) operational definition which was based upon measurements made from spectrographs." (Edwards 1981, pp. 539-540). The main results of VOT in most studies have however the same direction, irrespective of type of analysis or material, i.e. velars > alveolars > labials (see e.g. Lisker and Abramson 1964; Edwards 1981; Kewley-Port 1983; Cassidy and Harrington 1995).

The place of articulation distinctions do not show as clear cut differences as the voicing distinction. The results of the present study show duration differences of vowel onset time (i.e. burst), but none for the release spectral structure or for the post-consonant formant onset or formant transitions for each place of articulation. The vowel onset time (≅VOT), although widely acknowledged as a correlate of voicing distinctions, has also been found different for each place of articulation in studies of other languages (e.g. Kewley-Port 1983). Thus it seems that vowel onset time may function as a multi-cue for both voice and places of articulation distinctions. Although not all speakers used this acoustic parameter constantly in the present study, the results showed statistical significance. This simply implies that speakers employ different means to convey a phonetic category and may use alternative strategies to achieve their target. This is one of the numerous pieces of evidence for the one-to-many (and vice-versa many-to-one) correspondence between acoustic parameters and phonetic categories reported in the literature and the speech encoding-decoding processes. It should be noted that both segments of the burst, i.e. release and aspiration, show related structures of place of articulation correlations, an indication that they may function as one structural unit. The fact that there is a regular trade-off between occlusion, release and aspiration durations is an acoustic evidence of complementary distribution which indicates the involvement of vowel onset time in both voice and place of articulation distinctions.

The results of the release burst analysis in the present study are based on a single measurement at the onset of the release within a 10 ms frame window as a first attempt at an acoustic analysis of the acoustic correlates of the Greek stop consonants. Early work carried by Kenneth Stevens and Sheila Blumstein (Stevens and Blumstein 1978; Blumstein and Stevens 1979, 1980) emphasises the importance of the overall spectrum shape within a 25.6 ms time-window in the vicinity of the consonantal release for the place of articulation distinctions. This short-term spectrum includes both burst spectrum and vowel onset frequencies and is assumed to provide invariant acoustic cues for each place of articulation independent of the following vowel whereas the formant transitions provide secondary context dependent cues. The invariance of acoustic correlates for place of articulation has also been demonstrated by Kewley-Port (1983) who used dynamic running spectra. Her approach led to better results for place of articulation categorisation than the Stevens-Blumstein approach which was based on a static spectral analysis at a fixed time-point. Results of subsequent studies support the Stevens-Blumstein initial hypothesis in principle and provide acoustic evidence at the release burst spectrum (e.g. Lahiri *et al* 1984; Cassidy and Harrington 1995). The significance of the burst as an acoustic and perceptual correlate of place of articulation oppositions has also been recently demonstrated for French (Bonneau, Djezzar and Laprie 1993; Djezzar 1995). Djezzar (1995) reports that the release burst portions of 25 ms duration with no remaining vocalic segment could be correctly identified by 87% on an average for the labial, alveolar and velar voiceless stops in French.

The vowel formant onsets do not show any major structural differences except for the alveolar stops. The labial and velar stops are clustered quite close together and do not exceed, as a rule, the vowel formant difference limen which is within the range of 50-150 Hz (Mermelstein 1978). On the other hand, the alveolar stops are above the difference limen in comparison to both labials and velars for all speakers and this difference may be a classification cue for place of articulation distinctions. The formant transitions do not show structural differences except for the alveolar consonants. The pre-consonantal F1 is falling whereas the post-consonantal F1 is level-to-rising for all consonants. The pre-consonantal F2 is also falling for the labials and velars but level-to-rising for the alveolar one whereas the post-consonantal F2 is rising for the labial and velar stops but rather level for the alveolar stops. Thus, under the present experimental conditions, the formant transitions do not provide evidence for invariant acoustic correlates for *each* place of articulation in Greek. This does not concern however F3 transitions which are not analysed in the present study. Recent studies however report acoustic correlates of formant transitions which separate each place of articulation in English as well as other languages (cf. Sussman 1991; Sussman *et al* 1991; Sussman *et al* 1993). Thus Sussman and co-workers report "locus equations", i.e. straight regression lines formed by plotting F2 vowel onset transitions along the *y* axis and their corresponding vowel nucleus along the *x* axis, with high classification scores for the stop place of articulation categories. Cassidy and Harrington (1995) also report good classificatory scores (i.e. well above chance) for vowel F2 onset in the order b < d < g as well as the first three vowel formant slopes into the vowel nucleus.

The acoustic structure of F1 and F2 frequencies are in agreement and close to values with recent results on the Greek vowels (Botinis, Fourakis and Katsaiti 1995, and Fourakis, Botinis and Katsaiti 1999). The stressed vs. unstressed vowel values show a mirror image structure with higher F1 values and lower F2 values for the vowel of the unstressed syllable in comparison to the vowel of the stressed syllable. In the Botinis *et al.* (1995) study the prosodic condition of stress resulted in an expansion of the acoustic space whereas the effects of lack of stress were a compression in combination with a raised acoustic space. The effect of stress on the formant structure of the vowels is also evident in the present study. The formant transitions do not however show a similar structure in accordance with expansion vs. the compression effect of the stress condition. This is especially evident with the pre-consonantal F1 transitions which have lower values than the post-consonantal ones, contrary to the expectations of the stress condition effect. It may however be a complex interaction between stop consonant production and prosody and its function may be related with either category.

6. Conclusions

The present study is a first acoustic analysis of the Greek stop consonants and its results should be corroborated with additional evidence. Furthermore, these results lead to new questions which will require detailed analysis and a deeper insight into the acoustic structure of the Greek stop consonants. The voicing distinction is characterised by the voicing of the occlusion, the duration of the occlusion and the post-consonantal vowel onset timing. The voiced stops are fully voiced, have longer occlusion and early vowel onset whereas the voiceless stops may be partly voiced or voiceless, have shorter occlusion and delayed vowel onset combined with aspiration. The total duration of the voiced and voiceless stops is approximately the same and the duration of the occlusion and burst have a complementary distribution pattern. The place of articulation oppositions show vowel onset timing differences whereas the spectral shape of the release and the formant transitions do not provide sufficient acoustic evidence for *each* place of articulation. Neither voicing nor place of articulation oppositions have any effect on nucleus vowel frequencies.

A. Botinis, M. Fourakis, I. Prinou
Univ. of Athens, The Ohio State Univ., Univ. of Athens

REFERENCES

Abramson, A. S. & Lisker, L. 1995: "Voice timing for stop classification in conversational English". *XIII International Congress of Phonetic Sciences* 4, 128-131.

Blumstein, S. E. & Stevens, K. N. 1979: "Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants". *Journal of the Acoustic Society of America* 66, 1001-1017.

Blumstein, S. E. & Stevens, K. N. 1980: "Perceptual invariance and onset spectra for stop consonants in different vowel environments". *Journal of the Acoustic Society of America* 67, 648-662.

Bonneau, A., Djezzar, L. & Laprie, Y. 1993: "Perception of French stop bursts: implications for stop identification". *Eurospeech* 1, 693-696.

Botinis, A., Fourakis, M. & Katsaiti, M. 1995: "Acoustic characteristics of Greek vowels under different prosodic conditions". *XIII International Congress of Phonetic Sciences* 4, 404-407.

Cassidy, S. & Harrington, J. 1995: "The place of articulation distinction in voiced oral stops: evidence from burst spectra and formant transitions". *Phonetica* 52, 263-284.

Cooper, F. S., Delattre, P. L., Liberman, A. M., Borst, J. M. & Gerstman, L. G. 1952: "Some experiments on the perception of synthetic speech sounds". *Journal of the Acoustic Society of America* 24, 597-606.

Delattre, P., Liberman, A. M. & Cooper, F. S. 1955: "Acoustic loci and transitional cues for consonants". *Journal of the Acoustic Society of America* 27, 769-773.

Djezzar, L. 1995: "Effect of prior knowledge of the vowel on the perception of French stop bursts". *XIII International Congress of Phonetic Sciences* 2, 262-265.

Dorman, M. F., Studdert-Kennedy, M. & Raphael, L. 1977: "Stop consonant recognition: release bursts and formant transitions as functionally equivalent, context-dependent cues". *Perception and Psychophysics* 22, 109-122.

Edwards, T. 1981: "Multiple feature analysis of intervocalic English plosives". *Journal of the Acoustic Society of America* 69, 535-547.

Fant, G. 1960: *Acoustic theory of speech production* (The Hague: Mouton).

Fant, G. 1968: "Analysis and synthesis of speech processes". In B. Malmberg (ed.), *Manual of phonetics* (Amsterdam: North Holland-Publishing Company) 173-277.

Fant, G. 1973: "Stops in CV-syllables". In G. Fant (ed.), *Speech sounds and features* (Cambridge MA: MIT Press) 110 -139.

Fourakis, M., Botinis, A. & Katsaiti, M. 1999: "Acoustic characteristics of Greek vowels". *Phonetica* 56, 28-43.

Hazan, V. & Bo, S. 1995: "Individual variability in the perceptual weighting of cues to stop place and voicing contrasts". *XIII International Congress of Phonetic Sciences* 2, 506-509.

Henton, C., Ladefoged, P. & Maddieson, I. 1992: "Stops in the world's languages". *Phonetica* 49, 65-101.

Householder, F. 1964: "Three dreams of Modern Greek phonology". In R. Austerlitz (ed.), *Papers in memory of George C. Pappageotes* (issued as supplement to *Word* 20, 3) 17-27.

Jakobson, R. & Halle, M. 1956: *Fundamentals of language* (The Hague: Mouton).

Jakobson, R. & Halle, M. 1962: "Tenseness and laxness". In *Roman Jakobson, Selected writings.* I: Phonological Studies (The Hague: Mouton) 550-555.

Joseph, B. & Philippaki-Warburton, I. 1987: *Modern Greek* (Croom Helm).

Kewley-Port, D. 1982: "Time-varying features as correlates of place of articulation in stop consonants". *Journal of the Acoustic Society of America* 73, 322-335.

Kewley-Port, D. 1983: "Measurement of formant transitions in naturally produced stop consonant-vowel syllables". *Journal of the Acoustic Society of America* 72, 379-389.

Kewley-Port, D., Pisoni, D. & Studdert-Kennedy, M. 1983: "Perception of static and dynamic cues to place of articulation". *Journal of the Acoustic Society of America* 73, 1779-1793.

Koutsoudas, A. 1962: *Verb morphology in Modern Greek: A descriptive analysis* (Bloomington: Indiana University / The Hague: Mouton).

Lahiri, A., Gewirth, L. & Blumstein, S. 1984: "A reconsideration of acoustic invariance for place of articulation in diffuse stop consonants: Evidence from a cross-language study". *Journal of the Acoustic Society of America* 76, 391-404.

Larranaga, P., Lleo, C. & Prinz, M. 1995: "A cross-linguistic VOT analysis of early stop production: The development of the feature voicing". *XIII International Congress of Phonetic Sciences* 3, 440-443.

Liberman, A. M., Dellatre, P. C. & Cooper, F. S. 1952: "The role of selected stimulus-variables in the perception of the unvoiced consonants". *American Journal of Psychology* 65, 497-516.

Liberman, A. M., Dellatre, P. C., Cooper, F. S. & Gerstman, L. H. 1954: "The role of consonant-vowel transitions in the perception of the stop and nasal consonants". *Psychological Monographs* 68, 1-13.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. 1967: "Perception of the speech code". *Psychological Review* 74, 431-461.

Liberman, A. M. & Mattingly, I. G. 1985: "The motor theory of speech perception revised". *Cognition* 21, 1-36.

Lieberman, P. & Blumstein S. E. 1988: *Speech physiology, speech perception and acoustic phonetics* (Cambridge: Cambridge University Press).

Lisker, L. 1986: "'Voicing' in English: A catalogue of acoustic features signalling /b/ versus /p/ in trochees". *Language and Speech* 29, 3-11.

Lisker, L. & Abramson, A. S. 1964: "A cross-language study of voicing in initial stops: Acoustic measurements". *Word* 20, 384-422.

Lofqvist, A. 1980: "Interarticulator programming in stop production". *Journal of Phonetics* 8, 475-490.

Madiesson, I. 1984: *Patterns of sounds* (Cambridge: Cambridge University Press).

Mermelstein, P. 1978: "Difference limens for formant frequencies of steady-state and consonant-bound vowels". *Journal of the Acoustic Society of America* 63, 572-580.

Newton, B. 1961: "The rephonemicization of Modern Greek". *Lingua* 10, 275-284.

Newton, B. 1972: *The generative interpretation of dialect. A study of Modern Greek phonology* (Cambridge: Cambridge University Press).

Pagoni, S. 1993: *Modern Greek variation: A government phonology approach* (Ph.D. Thesis, University of London).

Repp, B. & Lin, H. 1989: "Acoustic properties and perception of stop consonant release transients". *Journal of the Acoustic Society of America* 85, 379-396.

Sonesson, B. 1970: "The functional anatomy of the speech organs". In B. Malmberg (ed.), *Manual of phonetics* (Amsterdam: North Holland-Publishing Company) 45-75.

Stevens, K. N. & Blumstein, S. E. 1978: "Invariant cues for place of articulation in stop consonants". *Journal of the Acoustic Society of America* 64, 1358-1368.

Studdert-Kennedy, M., Liberman, A. M., Cooper, F. S. & Harris K. S. 1970: "Motor theory of speech perception: A reply to Lane's critical review". *Psychological Review* 77, 234-249.

Sussman, H. M. 1991: "The representation of stop consonants in three-dimensional acoustic space". *Phonetica* 48, 18-31.

Sussman, H. M., McCaffrey, H. A & Matthews, S. A. 1991: "An investigation of locus equations as a source of relational invariance for stop place categorisation". *Journal of the Acoustic Society of America* 90, 1309-1325.

Sussman, H. M., Hoemeke, K. A. & Ahmed, F. S 1993: "A cross-linguistic investigation of locus equations as a phonetic descriptor of place of articulation". *Journal of the Acoustic Society of America* 90, 1309-1325.

Van den Berg, J. 1958: "Myoelastic-aerodynamic theory of voice production". *Journal of Speech and Hearing Research* 1, 227-244.

Umeda, N. 1977: "Consonant duration in American English". *Journal of the Acoustic Society of America* 61, 846-858.

Warburton, I. 1970: *On the verb in Modern Greek.* Language Science Monographs, 4 (Bloomington: Indiana University / The Hague: Mouton).

Zue, V. W. 1976: "Acoustic characteristics of stop consonants: A controlled study" (MIT: Technical report 523, Lincoln Laboratories).