

DUCCIO PICCARDI

## Sociophonetic factors of speakers' sex differences in Voice Onset Time: A Florentine case study

The paper shows the results of a sociophonetic analysis of the so-called *gorgia enfatica* (Castellani, 1960), i.e. the allophonic presence of voiceless aspirated plosives in strong positions attested in the vernacular variety of the city of Florence. 24 native speakers were involved in a production test (read speech) followed by a perceptual counterpart (*matched-guise* and open-ended interview). A statistically significant relationship between male speakers and longer Voice Onset Times emerged from our quantitative analysis of production. This distribution was recognized by the speakers, that evaluated the trait using comments fit for a tentative reconstruction of an *indexical field* (Eckert, 2008). Our data corroborate the need for a sociophonetic shift in the research methods concerning the relation between speakers' sex and VOT production (Oh, 2011).

*Key words:* Gorgia enfatica, Voice Onset Time, indexical field, gender, sociophonetics.

### 1. State of the art

#### 1.1 Voice Onset Time – speakers' sex patterns

Since the first definition of *Voice Onset Time* (VOT) in Lisker, Abramson (1964), this consonantal subsegmental feature has been studied in relation to many linguistic (voicing, speech rate, pitch, place of articulation, following vowel quality etc.) and social variables. In particular, the history of the studies about speakers' sex differences in VOT shows a recent interesting shift towards research paradigms focused on the observation of the trait as it is structured and motivated in a single local speech community, gradually abandoning explanations based on universal or physiological properties. Swartz (1992) was the first research dedicated to the analysis of VOT-sex patterns. The scholar noticed that female (F) English speakers produce voiceless plosives with significantly longer VOTs than their male (M) counterparts<sup>1</sup>. This pattern was later confirmed by Ryalls, Zipprer & Baldauff (1997), Whiteside, Irving (1997; 1998), Koenig (2000), Whiteside, Henry & Dobbin (2004), Robb, Gilbert & Lerman (2005) and Morris, McCrea & Herring (2008)<sup>2</sup>. It should be noted that not all the mentioned studies are comparable in terms of research

<sup>1</sup> The test consonants of Swartz (1992) were [t] and [d]. Data on this topic were also previously reported in Smith (1978) and Sweeting, Baken (1982). In addition to that, as Swartz himself mentions (Swartz 1992: 984), Molfese, Hess (1978) and Molfese, Molfese (1988) can be seen as examples of the beginning of a structured scientific tradition about the correlations between VOT and speakers' sex, focused in these first attempts on perception and not on production.

<sup>2</sup> A quantitative overview of the results presented in these studies can be found in Oh (2011: 60).

methods, neither they have a uniform evaluation of stylistic phenomena: for example, the status of speech rate remains at least controversial if we compare its non-significant effect in Swartz (1992) with the results presented in Morris *et al.* (2008), that, analyzing monosyllables, loses the statistical significance of the sex opposition. The apparent regularity of this VOT-sex pattern led to conclusions inferred from anatomical differences in speakers' phonatory apparatus, such as men's wider supraglottic space and women's shorter and stiffer vocal folds<sup>3</sup>. Observing the significant distance between female VOT values of voiceless and voiced plosives, Whiteside, Irving (1998) try to further add a broad stylistic and sociolinguistic argument: the authors hypothesize that the pattern is a consequence of an optimization of the voicing contrast in female speech, strictly correlated to a slower and clearer style<sup>4</sup>. This first explanatory tendency, that we may call the *physiological/universal line*, has proved insufficient as a result of a gradual diversification of the linguistic contexts taken into account. The first turning point was the observation of the evolution of VOT-sex patterns during the developmental age. Whiteside, Marshall (2001) found that male English speakers around the age of 9 have a dramatic increase in VOT lengths, not shown by their female counterparts and related to sex-determined physical growth. VOT values become adultlike around the age of 11: from the English male speakers' point of view, this is an inhibition of a physical potential, determined by sociophonetic factors<sup>5</sup>. Nissen, Fox (2009) focused on pre-pubescent children, pointing out that female greater values are established around the age of 5, before the growth mechanism explained by Whiteside, Marshall (2001). Since there is no appreciable sexual dimorphism in the vocal tract at the age of 5, the authors conclude that their results «may be based on male-female archetypes present in adult production patterns» (Nissen, Fox, 2009: 1376). The other fundamental turning point concerns the analysis of non-English-speaking communities. Oh (2011) was the first to find a statistically significant relation between male speakers and longer VOTs, through the observation of the aspirated voiceless plosives of Seoul Korean. The author links his findings to an ongoing change led by female speakers, that are substituting VOT length with f0 values in the phonological cues of the voiceless aspirated class<sup>6</sup>. Oh (2011) explicitly tries to undermine the *physiological/universal line*, promoting the birth of a new *sociophonetic line*: «[...] gender differences in VOT values index sociophonetic gender variations, and [...] sociophonetic patterns can differ across languages or dialectal group» (Oh, 2011:

<sup>3</sup> Both of these features grant male speakers a facilitation in «the formation of sub and supraglottic pressure differences and therefore vocal fold vibration» (Oh, 2011: 60). See Smith (1978), Pickett (1980), Swartz (1992), Titze (1994), Whiteside, Irving (1997; 1998), Koenig (2000) for further discussion on this point.

<sup>4</sup> This claim seems to reflect the results of Scharf, Masur (2002) based on German speakers and those of Peng, Chen & Lee (2014) (Sixian Hakka and Taiwanese Mandarin); however, caution should be taken in the indiscriminate application of this criterion to other cases of VOT differences as cues of phonemic distinction. Alzoubi (2016) shows for example greater emphatic-plain contrast in male Jordanian Arabic speakers (involving VOT length) than in their female counterparts.

<sup>5</sup> A later similar study (Whiteside *et al.*, 2004), does not show the same tendency for male speakers. The idea has been recently rediscussed in Yu, De Nil & Pang (2015), with particular attention to its sociophonetic implications.

<sup>6</sup> This change has been confirmed by Kang (2014) and further extended to the affricates by Lee (2016).

65)<sup>7</sup>. In other words, researchers should evaluate the sociophonetic potential of VOT in the examined speech communities before interpreting distributional data as an outcome of the sexual dimorphism of the phonatory apparatus or common behavioral patterns. In recent years, many scholars added elements to our diatopic knowledge about this correlation. Abudaljuh (2011) found no statistical significance for his F > M VOT results regarding the emphatic plosives of Jordanian Arabic. Swedish does not seem to have any significant VOT-sex pattern, as shown in Lundeborg Hammarström, Larsson, Wiman & McAllister (2012). Li (2013) (Mandarin Chinese of Songyuan) loses the statistical significance of his F > M long-lag VOT tendency after controlling the effect of speech rate, but maintains under the same conditions a strong M > F pattern for short-lag VOTs. Cheng (2013) reports very slight sex-related differences for VOTs in Sixian Hakka, showing non-significant F > M for the aspirated series and M > F for the unaspirated one. These tendencies for Sixian Hakka acquire statistical significance in Peng *et al.* (2014), comparable to a similar trend found for the Mandarin of Taiwan. Madhu, Mahendra & Sreedevi (2014) add to this picture some values taken from Telugu plosives. In this Dravidian language, the authors find a general M > F pattern, even if significant only for few plosive types. In Brinca, Araújo, Nogueira & Gil (2016), we find data about a Romance language, Portuguese, in which pre-adolescent children do not show sex-related differences in VOT length. Coming to the Italian dialectal area, only some Calabrese varieties<sup>8</sup> are covered in relation to our topic. The partial results of Romito and Ciardullo research project in San Giovanni in Fiore (Romito, Ciardullo & Tarasi, 2015) notice a slight, non-significant F > M tendency. A non-significant M > F trend was found in Lamezia Terme by Nodari, 2015, even if a strong interaction between the sex variable and the selected class variable (*orientamento verso la scuola*, “scholastic proficiency”) reconstructs a suggestive example of sex-prestige pattern, as the lowest-class male speakers produce the longest VOT lengths, interpreted as a sub-standard trait. As shown by this overview, our knowledge of VOT-sex patterns as they are structured in various languages of the world dramatically increased since Oh (2011), that remains *de facto* the only study showing and motivating a strong M > F tendency for voiceless plosives. Unfortunately, few researchers until now decided to embrace Oh's suggestion about following a *sociophonetic line*, reporting instead unexplained distributional data<sup>9</sup>. In order to tackle one of the biggest challenge of sociophonetics, the distinction between socially meaningful traits and biologically determined ones (see Calamai, 2015: 71), we should

<sup>7</sup> Before the fundamental Oh (2011) few other authors tried to exit the anglophone area without finding statistically significant results (e.g. Bijankhan, Nourbakhsh, 2009, for Persian). Interestingly enough, Oh (2011) does not quote the results of Han (2010), that show English-like VOT-sex patterns in the southeastern Korean city of Daegu. We can consider the comparison between the Daegu and the Seoul distribution as one of the few existing proofs of Oh's claim about the possibility of sociophonetic interdialectal differences in this correlation.

<sup>8</sup> In Calabrese dialects aspirated voiceless plosives are a remarkable local trait (e.g., Soriano, 1996).

<sup>9</sup> Exceptions can be found in Abudaljuh (2011) and Nodari (2015). However, these authors make sociophonetic inferences from their distributional results, without anchoring production to speakers' perception and evaluation. Cheng (2013: 215) weights up the importance of the observed patterns in explaining some acquisitional trends of foreign brides coming to Hakka-speaking communities, but the topic is no further discussed.

try to answer questions like: what type of indexicality VOT values hold? What is speakers' evaluative perception of the trait? Is there a structurally motivated reason to the selection of a specific VOT-sex pattern? We kept in mind this goal during the analysis of an Italian dialectal trait, the so-called Florentine *gorgia enfatica*.

## 1.2 The *gorgia enfatica*

Castellani (1960) named *gorgia enfatica* the allophonic presence of voiceless aspirated stops ([k<sup>h</sup>, t<sup>h</sup>, p<sup>h</sup>]) in strong position in the vernacular variety spoken in Florence. The trait was firstly noticed during Scheuermeier's fieldwork and then reported (with a dated and now ambiguous transcription system) in the *Linguistic and Ethnographic Atlas of Italy and Southern Switzerland* (AIS). Giacomelli (1934) tried to further investigate the Tuscan chief-town in order to find confirmation of the allophones. The researcher not only recognized the phenomenon, but also collected an interesting metalinguistic comment about it interviewing a group of native speakers (Giacomelli, 1934: 195-200). Moving to the exegetical hypotheses, two divergent theories about the *gorgia enfatica* have been formulated during the second half of the last century. In Arrigo Castellani's (1952; 1960) substratist view, the *gorgia enfatica* can be found only in Florence and in some areas of the Mugello, and represents the last observable remnant of the Etruscan phonological series of voiceless aspirated stops, reduced in weak positions to fricative allophones (the *gorgia*). However, during the following years, similar consonantal strengthening processes were found not just in other Tuscan areas, but also in the entire Italian linguistic domain (see Izzo, 1972). This led to a series of interpretations that saw the *gorgia enfatica* as not significantly linked to the dialects of Tuscany or to Florence, but generated by prosody (i.e., hyperarticulation in Izzo, 1972; emotional speech in Giannelli, 1976) or universal articulatory gestures (Giannelli, Savoia, 1978; 1979-1980, Giannelli, 1983). These two lines are supported without distributional experimental studies about the phenomenon; moreover, following the anti-substratist exegetical path does not explain the peculiar data registered in the AIS, nor Giacomelli's verification of the trait. This clash between local significance and general interpretations<sup>10</sup>, somehow reflecting what we said about VOT-sex patterns, seemed to us the ideal research field to test the exegetical potential of sociophonetic methodologies.

## 2. *Methods of the sociophonetic investigation*<sup>11</sup>

### 2.1 Participants, setting and structure of the protocol

24 speakers, all born and living in Florence, were recruited for this research. Most of the participants knew us from a long time; if additional individuals belonging to specific so-

<sup>10</sup> From the point of view of traditional Italian dialectology, a substratum hypothesis loses reliability if the trait under inquiry can also be found outside the geographical area of its supposed linguistic substratum (i.e. the so-called *chorographic* requirement).

<sup>11</sup> What we present here is a thematic excerpt and in-depth analysis of the results reported in Piccardi, (in press). In the op. cit. the reader will find reflections about the historical depth of the trait with an attempt of disambiguation of the problematic use of the term "emphasis". In this sense, a prosodic

cial classes were required, we asked the already agreeing participants to search for plausible candidates<sup>12</sup>. All the speakers involved were aware that they were contributing to a dialectological research about the variety of their city; however, we did not provide further information until the final debriefing (see below). Following the methods of one of the few available sociolinguistic studies on the normal *gorgia*, Cravens, Giannelli (1995), we selected three social variables, M vs F (sex variable), I vs II vs III (age variable, adjusted to reflect a generational criterion<sup>13</sup>) and G vs NG (Graduates vs NonGraduates, class variable<sup>14</sup>). We tested two subjects per each possible combination. The survey was conducted contacting two speakers per session. These dyads were composed by socially homogeneous individuals, already acquainted with each other. In order to minimize auditor effects<sup>15</sup>, the only other person on the scene of the interviews was the researcher. The tests were performed in quiet rooms of private habitations. We built for our purposes a three-step mixed methods protocol based on the integration of quantitative analysis of production with qualitative evaluation of speakers' perception. The three phases of the protocol were presented to our participants during a single session, following a fixed order. The duration of each session was approx. 45 mins. The general structure of the protocol was devised with the aim of triangulating our data through *significance enhancement* (Collins *et al.*, 2006: 83-87), i.e., the potential mutual optimization of the understanding of quantitative and qualitative results, and the discovery of causal relationships through the attribution of meaning to quantities.

## 2.2 Production task

*Materials.* We prepared a session of read speech consisting in a story-telling task and 60 sentences built around a selection of target words containing /k, t, p/ in the following "strong" contexts: initial (*C-*), syntactic doubling (RF, *Raddoppiamento Fonosintattico*<sup>16</sup>),

---

variable of emphasis was tested explicitly suggesting an emotive utterance of two of the four frames presented (see below). This aspect of the structure of our research should be taken into account in the comparison between our production data and other sources. VOT lengths discriminated by the prosodic variable can be found in our mentioned paper.

<sup>12</sup> This procedure is called the *friend of a friend method* (Milroy, 1987: 52-56), and is usually considered useful to create a privileged position for the observation of naturalistic behaviors.

<sup>13</sup> The cut-off ages selected in Cravens, Giannelli (1995) were 15-25, 35-45 and above 65. Means and standard deviations of our revised categories are as follows: I (25; 1); II (47,5; 5,17); III (73,5; 7,34).

<sup>14</sup> About the class variable, we decided to reformulate the binary partition chosen in the op. cit. (blue collars vs white collars) using Hudson's idea of "sophistication" (Hudson, 1996: 198-199). Hudson proposed a macrosocial division based on the stereotypes of "rough" and "sophisticated" people. This division has predictable consequences on the sex-prestige pattern, and is appreciably practical since it excludes census, focusing on cultural and occupational traits of the speakers. We decided to not introduce other layers of social distinction to maintain a degree of comparability with the previous (AIS etc.) twofold distributions. In our research, a speaker was considered "sophisticated" if a) holding a university degree, and b) having an occupation that requires the aforementioned degree. Conversely, a "rough" speaker was selected if a) not holding a university degree, and b) having an occupation that does not require any kind of degree. Hybrid cases were excluded from this bipartition.

<sup>15</sup> This terminology was introduced by Bell (1984) to describe the effects on style-shifting of the presence on the scene of a conversation of other interlocutors besides the addressee.

<sup>16</sup> See, e.g. Lopporcaro (1997: 41-51).

gemination (-C-), postconsonantal (-CC-) and postsibilant (-sC)<sup>17</sup>. Each target-word was repeated in 4 different frames, presented to the speakers written in a morphosyntactically marked variety of Regional Tuscan Italian. For this first experimental verification of the *gorgia enfatica*, the words were chosen from the available bibliography (AIS; Castellani, 1960) reporting impressionistic data. This study is thus not concerned with potentially conditioning factors, such as word length, frequency, prosodic structure and the quality of the vowels following the plosives<sup>18</sup>. Anyway, the words were not placed in prepausal positions, so as to rule out the effects of contextual lengthening.

*Procedure.* The dyads were first invited to have an informal conversation in the room of the interview (approx. 20 mins.). We tried to limit our direct verbal interactions in the conversations: in fact, this phase was devised with the aim of mitigating the participants' *attention to speech* (see, e.g. Labov, 1984), accustoming them to have spontaneous discussions while we were observing the scene. Then, a sheet of paper with the texts of the short story and the 60 sentences was given to each participant. The speakers were asked to read the written contents addressing the other participant, using an informal tone. They were allowed to read silently the proposed excerpts before pronouncing them. One single production was digitally recorded for each speaker, using a SONY ICD-UX533.

*Analysis.* In order to increase the number of our tokens, all voiceless plosives in strong positions elicited during the production phase were analyzed in PRAAT 6.0.14 through manual annotation of spectrograms. Our corpus consisted of 1041 /k/, 2287 /t/ and 738 /p/. Only VOTs preceded by a release burst were considered valid (Marotta 2004)<sup>19</sup>. Measurements were taken from the release burst to the onset of periodicity in the acoustic waveform (Francis, Ciocca & Yu, 2003). The reported results will not consider potential effects of speech rate on VOT lengths<sup>20</sup>. A series of one-way ANOVA statistical tests were performed in Matlab R2016a<sup>21</sup>, with the aim of observing the effects of sex (df = 1), class (df = 1) and age (df = 2) comparing the VOT means for each phoneme in each linguistic context. We controlled the family-wise inflation of type I error adjusting the  $\alpha$  significance level with a Bonferroni correction, dividing it by the number of performed tests ( $n$ ), i.e.  $\alpha/n = 0,05/45 = 0,001$ .

<sup>17</sup> This last context was discriminated from the generic postconsonantal one to better evaluate the potential inhibition on the production of long-lag VOTs triggered by assimilations of manner and laryngeal features (Izzo, 1972: 107).

<sup>18</sup> For this last aspect, some observations about the lengthening effect of the following high vowels have been presented in Piccardi (in press). The tested words are listed below; <<sup>h</sup>> stands for the position of the aspirations discussed in our sources: *ch<sup>b</sup>e* ("that"), *c<sup>b</sup>ane* ([i'k:<sup>h</sup>ane], "the dog"), *tabac<sup>b</sup>o* ("tobacco"), *merc<sup>b</sup>ato* ("market"), *nasc<sup>b</sup>ondino* ("hide-and-seeK"), *t<sup>b</sup>u* ("you"), *sott<sup>b</sup>errare* ("to bury"), *t<sup>h</sup>abernacolo* ([i t:<sup>h</sup>aber'naholo], "the aedicule"), *sol<sup>b</sup>anto* ("only"), *st<sup>b</sup>azione* ("station"), *p<sup>b</sup>ettinare* ("to comb"), *p<sup>b</sup>aese* ([i p:<sup>h</sup>a'eze], "the village"), *cepp<sup>b</sup>o* ("stump"), *un p<sup>b</sup>o'* ("a bit"), *asp<sup>b</sup>ettare* ("to wait").

<sup>19</sup> A small subset of plosives showed indeed a short release phase without an audible release burst. These were classified as belonging to a different phonetic category (unreleased plosives; see e.g. Mioni, 2001: 39) and thus excluded from our computation.

<sup>20</sup> We thank an anonymous reviewer for providing us with the reference to a recent paper discussing the need for speech rate normalization in VOT measurements (Nakai, Scobbie, 2016). We intend to take part to this ongoing debate in a separate publication.

<sup>21</sup> We deeply thank Dr. Federico Becattini (University of Florence) for the assistance in this part of the research.

### 2.3 Perception tasks<sup>22</sup>

*Materials.* Three pairs of short acoustic stimuli (approx. 5 secs. each), one per plosive point of articulation, were prepared for an informal matched-guise session (Lambert, 1967), recording the voice of the researcher (native speaker of the Florentine dialect) with the same, above stated equipment. The variables were situated in new (i.e., not previously presented) lexical items, with the aim of dismissing the possibility for eventual production-perception correlations of being constrained at the word level. The phrases had a content typologically homogeneous to the tone of the excerpts used for the production phase, i.e. light-hearted and conversational. First, three sentences were spontaneously read by the researcher while not producing the *gorgia enfatica*. These three excerpts were labelled as non-aspirating guise, also serving as a VOT durational baseline ([k]: 28 ms., [t]: 20 ms., [p]: 17 ms.) for the production of aspirating counterparts. The aspirating guise was enacted trying to replicate the sentences using the *gorgia enfatica*, aiming to add approx. 30 ms.<sup>23</sup> to the VOT baselines while maintaining a comparable intonational pattern. The best VOT candidates ([k<sup>h</sup>]: 62 ms., [t<sup>h</sup>]: 51 ms., [p<sup>h</sup>]: 52 ms.) were selected checking the spectrograms in PRAAT. The last part of our three-step protocol consisted in an open-ended interview (Marcato, Ursini & Politi, 1974: 28-30) about the metalinguistic competence of our participants. A topic guide was previously written and printed on a sheet of paper, readable only by the researcher. The guide was structured so as to progressively narrow the focus of the discussion, starting from topics of very broad linguistic interests and ending with a metalinguistic commentary on the trait under scrutiny, not without a direct confrontation with our research questions<sup>24</sup>. Six main topics of discussion were tackled during the interviews: 1) local terminology for non-standard speech varieties; 2) situations considered suitable for the above mentioned varieties; 3) non-standard traits emerging in the above mentioned situations; 4) local terminology for the most salient of the above mentioned traits, i.e. the *gorgia*; 5) linguistic contexts of the above mentioned trait; 6) [contrastively] metalinguistic discussion about the phenomena occurring in the specular contexts, i.e. the *gorgia enfatica*.

*Procedure.* The recorded excerpts were reproduced in a media player through the integrated speakers of a laptop (Asus SonicMaster). The six stimuli were presented as pairs of different guises; however, no internal order of reproduction was fixed. The rhythm of the test was dictated by the demands of the hearers, and multiple reproductions of the materials were permitted. Two open-ended questions were asked to the participants: "Are there any differences between the first and the second recording, and, if yes, what do they

<sup>22</sup> This phase of the inquiry has been inspired by the qualitative component of Campbell-Kibler's (e.g. 2007: 33-39) classic matched-guise protocol, reorganizing her open-ended questions, explicit identification of the changing variables and guise opening.

<sup>23</sup> We choose this value in the 10-40 ms. of the *Just Noticeable Differences* (see Calamai, Ricci, 2005: 73-74) in terms of length considering that «the subjects are much more sensitive to changes in vowel duration than to changes in consonant duration» (Huggins, 1972: 1270).

<sup>24</sup> This top-down procedure is usually called *funneling*, «an indirect way of approaching the heart of the matter, and may be used as a way of manipulating the respondents' degree of awareness of the object of inquiry» (Kristiansen, 2010: 538), enhancing it.

consist in?”<sup>25</sup>; “What can you tell me about speaker X and speaker Y?”. Participants’ answers were recorded and summarized taking written notes. The same was done during the open-ended interview. The last, more important point of the discussion was presented in three consecutive phases. First, participants were invited to reproduce the *gorgia enfatica*, instructing them to emit a puff of air after the normal consonants. Then, the previously heard guises were opened, i.e. the participants were made aware that they were evaluating the same person enacting two different personae<sup>26</sup>. Finally, they were asked to directly comment on our research questions. More information on the research were provided to those who manifested interest.

Table 1 - VOT values of the Male (M) and Female (F) classes: Means ( $\mu$ ) and Standard Deviations ( $\sigma$ ) calculated in ms. Statistical significant values after the Bonferroni correction can be found in the black frames

	$\mu M$	$\sigma M$	$\mu F$	$\sigma F$
<i>k-</i>	39,1	10,1	36,1	13,6
<i>RFk</i>	39,2	12,6	37,3	18,8
<i>k:</i>	38	13,4	39,4	14,8
<i>Ck</i>	34,4	10,3	30,3	10,1
<i>Sk</i>	29,6	10,6	27,7	9,8
<i>t-</i>	29,2	9,7	23,9	9,5
<i>RFt</i>	25,2	10,4	20,4	8,8
<i>t:</i>	27,3	10,5	23,3	12,4
<i>Ct</i>	23	8,3	19,5	7,8
<i>St</i>	19,7	6,6	17,9	5,6
<i>p-</i>	14,9	5,8	12,1	6,1
<i>RFp</i>	16,5	7,5	15,6	11,3
<i>p:</i>	19	7,3	15,4	7,7
<i>Cp</i>	18,2	7,6	19	9,4
<i>Sp</i>	13,4	5,5	10,4	5,4

<sup>25</sup> This short AX discrimination task could be seen as based on a rhetorical question, since the stimuli were created *ad hoc* to be discriminated. The discrimination was instrumental to a) rule out the presence of ideological processes inhibiting participants’ will to point out differences (i.e. *erasures*; Irvine, Gal, 2000); b) focus participants’ attention on the analyzed variable.

<sup>26</sup> The aim of guise opening is to retrieve impressions on the variable as an individual stylistic resource (Soukup, 2013).



### 3. Results

#### 3.1 The quantitative analysis of production

A general M > F tendency is observable throughout our considered contexts. Some of them hold statistical significance even after the Bonferroni correction (see Tables 1-3): *-Ck-* (F = 11,84, p = 0,0006); *t-* (F = 16,52, p = 0,00007), *-RFt-* (F = 17,38, p = 0,00004), *-t:-* (F = 17,61, p = 0,00003), *-Ct-* (F = 41,64, p = 1,97<sup>-10</sup>); *-sp-* (F = 11, p = 0,001)<sup>27</sup>.

Table 2 - VOT values of the NonGraduated (NG) and Graduated (G) classes: Means ( $\mu$ ) and Standard Deviations ( $\sigma$ ) calculated in ms. Statistical significant values after the Bonferroni correction can be found in the black frames

	$\mu$ NG	$\sigma$ NG	$\mu$ G	$\sigma$ G
<i>k-</i>	40,1	11,6	35,1	12,1
<i>RFk</i>	41,3	17,2	35,1	13,4
<i>k:</i>	42,5	14,2	35,6	13,3
<i>Ck</i>	34,3	10,7	30,5	9,8
<i>Sk</i>	30,6	10,8	26,1	9,1
<i>t-</i>	30,3	10,5	23	7,9
<i>RFt</i>	25,1	10,9	20	7,6
<i>t:</i>	27,2	12,7	23	10
<i>Ct</i>	23,6	8,7	19,3	7,4
<i>St</i>	19	6,2	18,5	6,1
<i>p-</i>	14,2	6,5	12,7	5,6
<i>RFp</i>	16,6	10,4	15,5	8,4
<i>p:</i>	16	8,9	18,4	6,1
<i>Cp</i>	19	8,4	18,3	8,8
<i>Sp</i>	12,9	6,7	10,9	4,1

Our data confirm the “vernacular” distribution suggested by the AIS, Giacomelli and Castellani. A constant pattern NG > G is evident in our corpus, with widespread statistical significance: *k-* (F = 11,58, p = 0,0007), *-k:-* (F = 11,81, p = 0,0007), *-Ck-* (F = 9,95, p = 0,001); *t-* (F = 30,48, p = 1,19<sup>-7</sup>), *-RFt-* (F = 16,68, p = 5,76<sup>-3</sup>),

<sup>27</sup> *k-* (F = 3,99, p = 0,04), *-st-* (F = 8,69, p = 0,003), *p-* (F = 6,24, p = 0,01) and *-p:-* (F = 8,91, p = 0,003) lost their significance after the adjustment.

-t- (F = 26,49, p = 3,61<sup>-7</sup>), -Ct- (F = 49,26, p = 5,05<sup>-12</sup>)<sup>28</sup>. Few contexts proved significant for the age variable, none after setting  $\alpha = 0,001$ ; moreover, no general tendencies can be observed<sup>29</sup>. This led us to search for particularly relevant oppositions between the productions of subsets collocated in single age categories, developing differences on the sex and class axes. This seemed to be the case: three radical oppositions emerge from one-way ANOVA tests comparing the means of all VOT values between specific social subsets, i.e. M I G > F I G (F = 94,92, p = 5,01<sup>-21</sup>), M II NG > M II G (F = 94,79, p = 6,03<sup>-21</sup>) and M III NG > F III NG (F = 67,88, p = 9,59<sup>-16</sup>).

### 3.2 The qualitative analysis of perception<sup>30</sup>

#### 3.2.1 The matched-guise test

Every speaker was able to perfectly recognize the segmental place of variation and the modalities of the variation itself. Unfortunately, our 30 ms. VOT distinctions were often considered an unnatural emphasisization of “florentinity” or a clumsy imitation of the Florentine dialect made by a foreigner (see below). Nevertheless, 13/24 speakers recognized [k<sup>h</sup>] as “more Florentine”; similar comments were elicited by 10/24 speakers for [t<sup>h</sup>] and by 2/24 speakers for [p<sup>h</sup>]. This scale of proneness to identify consonantal allophones reflects the metalinguistic consciousness about the targets of the weakening process of the normal *gorgia*<sup>31</sup>. Moreover, as it is shown by the reported means, bilabial allophones with a perceivable aspirated release phase are indeed rare. The social coordinates attributed to the aspirated allophones, if defined as “more Florentine”, were uniform and consistent with the distribution of production values. The *gorgia enfatica* was considered manlier and more appropriate in informal situations. Interestingly enough, only male speakers specified those situations as funny conversations among friends. Three speakers evoked very specific images related to the utterance of aspirated allophones. F III NG 2 remembered the face of his grandfather, slightly moved. She reproduced his voice replicating the aspiration process on a new lexical item, [ink<sup>h</sup>an'tina] (“in the cellar”). F II NG 2 thought about an old husband tiredly speaking to his wife. She also described the setting of her image, a typical Tuscan living room. M III G 2 identified the aspirat-

<sup>28</sup> -RFk- (F = 8,49, p = 0,003), -sk- (F = 5,70, p = 0,01) and -sp- (F = 4,62, p = 0,03) lost their significance after the adjustment. /t/ aspirated allophones seem to hold a considerable degree of potential social distinctiveness (Hudson, 1996:249). Along this line, we noticed some reconsiderations in the distinctiveness view of aspirated allophones after hearing the /t/ ones.

<sup>29</sup> Singh, Keshet, Gencaga & Ray (2016) come to a similar conclusion with the analysis of a large corpus of 630 American English speakers. The absence of biologically determined VOT-age pattern does not imply the impossibility for VOT values to index a social category linked to age criteria.

<sup>30</sup> In the next two sections, we will briefly comment only the perceptual results instrumental in explaining the retrieved VOT-sex pattern. For further details, see Piccardi (in press).

<sup>31</sup> The exact quantities of the reactions are reported here for the sole purpose of exemplifying the different impacts on the metalinguistic consciousness of the participants caused by the three plosives. The proposed qualitative data should be interpreted as a hint towards the coexistence of exaggeration and dialectality in a single trait.

ing voice as a father speaking in an authoritarian way to his son. Upon our request to mentally try to switch the sex of the imagined figures (i.e., grandmother, wife, mother to daughter), the speakers firmly denied the possibility: in their sensibilities, the aspirated allophones were linked to an emotional universe connoted as virile. Finally, older speakers were more inclined to identify the aspirated allophones as part of the Florentine diasystem, followed by the young participants and the members of the second age group (III > I > II).

Table 3 - *VOT values of the First (I), Second (II) and Third (III) age classes: Means ( $\mu$ ) and Standard Deviations ( $\sigma$ ) calculated in ms*

	$\mu I$	$\sigma I$	$\mu II$	$\sigma II$	$\mu III$	$\sigma III$
<i>k-</i>	35,2	12	39,3	13	38,3	11,1
<i>RFk</i>	41,9	19,7	37,8	11,7	34,3	12,4
<i>k:</i>	40,1	14,8	40,3	13,5	36,6	13,8
<i>Ck</i>	33	11,5	32,4	9,9	32	10
<i>Sk</i>	29	11,3	26,8	9,6	29	9,7
<i>t-</i>	27,1	11,3	28,1	10,9	24,2	7,1
<i>RFt</i>	22,1	9,9	24	11,2	22,3	8,2
<i>t:</i>	26,1	12,8	26,4	11,9	23,4	9,7
<i>Ct</i>	21,7	8,4	20,8	8,5	21,1	7,9
<i>St</i>	18,2	6,9	19,9	6	18,1	5,3
<i>p-</i>	14	5,7	14,4	7,6	12	4,5
<i>RFp</i>	16,7	11,3	15	7,9	12,7	7,9
<i>p:</i>	18	7,7	17,5	7,7	15,8	7,6
<i>Cp</i>	20,9	8,9	17,1	7,1	18,1	9,2
<i>Sp</i>	12,6	6,7	11,8	5,3	11,2	4,6

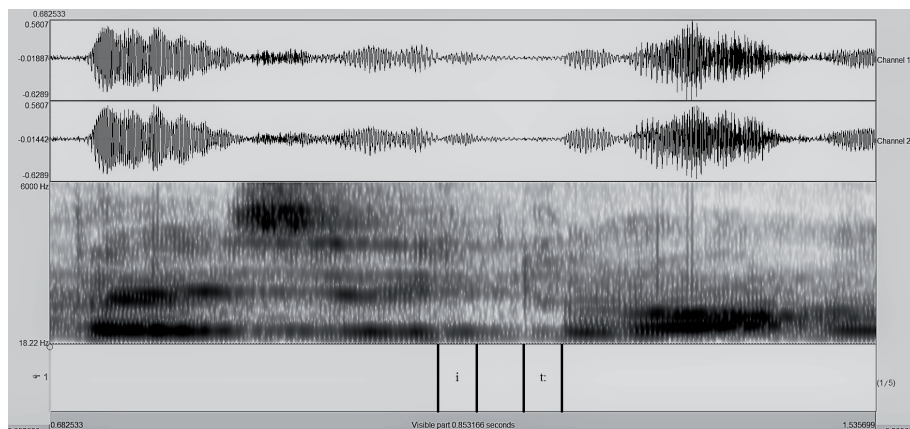
### 3.2.2 The open-ended interview

This order was not reflected by the metalinguistic competence emerged from our open-ended interviews. In our results, metalinguistic competence increased with the level of speakers' education (G > NG) and decreased with age (I > II > III)<sup>32</sup>. Probably the most interesting recurrent element emerged from our interviews is the referential ambiguity in the popular use of the word *aspirazione* ("aspiration"). *Aspirazione* has been used to refer to the Tuscan spirantization phenomenon since the XVI century, and is still today much more popularly widespread than the term *gorgia*. As Izzo, 1972 pointed out, this use of the term can no longer be considered acceptable in scientific contexts, for its potentially confusing ambiguity with proper aspiration processes (i.e., IPA [h]). Following the informal nature of the open-ended interview model, we used with our speakers *aspirazione* as a synonym of *gorgia*.

<sup>32</sup> These impressions were also anchored to a quantification of the traits suggested during our third point "What are the elements of your speech that you use in the situations you previously considered fitting for the dialect?". No significant differences were observed between M and F speakers.

However, we soon noticed that the speakers spontaneously extended the referential field of the word to the observation of real aspirates in strong positions, i.e. the *gorgia enfatica*. These unexpected extensions happened 11 times with different speakers, that introduced new lexical items in the conversation as examples of what they meant with *aspirazione* (*gorgia enfatica*; see Piccardi, in press, for further details). We had to infer that for our speakers *aspirazione* did not refer to a precise phonetic identity, but to an acoustical impression of a continuous sound that partially or totally places itself in the slot of a St. Italian plosive<sup>33</sup>. Coming to the last point of our interview, the explicit metalinguistic reflection about *the gorgia enfatica* with a guided pronunciation of the phenomenon, all the members of I were instantly capable to comment and identify the trait<sup>34</sup>. As we already mentioned, this ability decreased with age (II: 5/8; III: 3/8).

Figure 1 - Paese di tuo nonno (“village of your grandfather”) pronounced with RF by MI  
NG 1. VOT length of [tʰ]: 41 ms



The main social coordinate attributed to the trait was “typical of male speakers”, emerged in 12 interviews. Among these, 5 female speakers further specified that they heard the trait uttered by rustic, rude men (NG). Six interviews also presented a secondary attribution to III, class considered the depositary of archaic, authentic and rustic linguistic traits<sup>35</sup>; attributions to I were implicit in some diaphasic con-

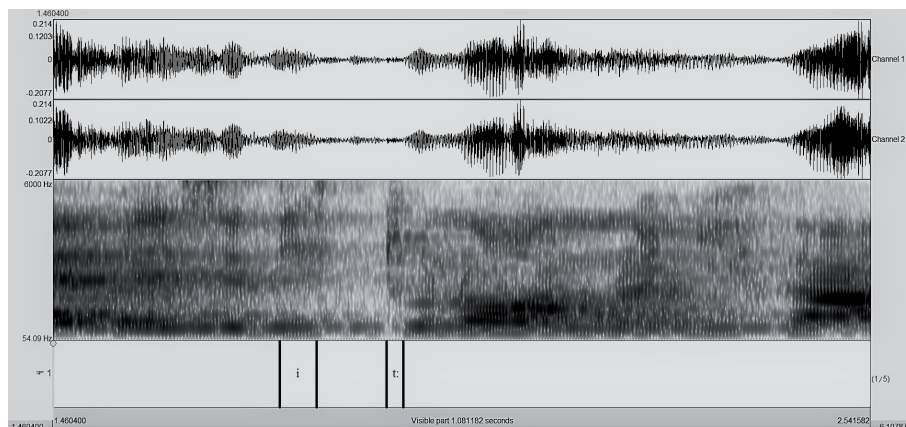
<sup>33</sup> The equivalence is plausible from an acoustic point of view, from the moment that [h] (the most typical and recognized realization of the Florentine *gorgia*) and [ʰ] share the same properties (Stevens, 1999: 451).

<sup>34</sup> All the data and impressions presented below were gathered during the phase before the last overt discussion on our research questions.

<sup>35</sup> An anonymous referee advanced the intriguing hypothesis that speech rate could be a determining factor in this association. From the moment that a) speech rate reportedly decreases with aging (Yuan, Liberman & Cieri, 2006); b) tempo has a strong effect on the perception of speakers’ age (Skoog Waller, Eriksson & Sörqvist, 2015); c) VOT length can increase as speech rate decreases (Summerfield, 1975); hearers could possibly link long VOTs with their occurrences in slow speech styles, and thus attribute them to old speakers. At this point of the research, we have no concrete elements to evaluate this lead. During the open-ended interview, our participants were reflecting on their own guided pro-

siderations. *Gorgia enfatica* was in fact perceived as a trait used in informal situations among friends in 8 interviews. At this point, any references to members of the family disappeared. Of those 8 interviews, 6 were with male speakers and the other 2 (young women) explicitly recognized the trait as used in groups of male-only friends. Finally, we found a M I tendency to celebrate the trait (“a sign of Florentine pride”: M I N G 1; “a way to express our peculiar way of being Florentine”: M I G 1) opposed to a F I one to firmly reject it (“yuk! It’s only for boys”: F I N G 2; “it’s just for rude old countrymen!”: F I G 2).

Figure 2 - *Paese di tuo nonno* (“village of your grandfather”) pronounced with RF by F I G 1.  
VOT length of [t]: 22 ms



#### 4. From sex to gender

In our research, we saw a degree of consistency between the distribution of long-lag VOTs revealed by production data and the evaluation suggested by our speakers. We can now try to move to an exegetical section and see if there are any socially structured reasons for this pattern, overriding supposed biological factors, and if aspiration can be invested by a peculiar set of social meanings. In order to dwell on these aspects, we should interpret Oh’s reference to the speaker *gender* not just as an unmarked synonym of biologically determined *sex*, but in the full sense of «social elaboration of biological sex», «not something we are born with, and not something we have, but something we do» (Eckert, McConnell-Ginet, 2003: 10). We will therefore shift from now on the meaning of our M and F from predetermined tags of social categories to social constructs composed by variable bundles of meanings (Kiesling, 2013: 461).

---

duction of long VOTs; moreover, the connection with speakers’ age was never elicited independently from the ones with speakers’ sex and conversational context. In our first approximation, we feel more confident in considering the reference to old speakers as a function of the general sex-prestige pattern.

#### 4.1 Explaining the roleplay

Somehow replicating the experience described in Giacomelli (1934), a key element to stimulate metalinguistic reflection about our analyzed trait was the phase of guided production of it. The importance of production for evaluation is one of the most explicit prerequisites (Iannàccaro, 2002: 137-138) for the insertion of a linguistic trait in the Labovian class of *markers*, composed by elements «not at the same high level of social awareness [of stereotypes], but show[ing] consistent stylistic and social stratification» (Labov, 1994: 78). A median position in the three-step Labovian scale of awareness does not fully explain the articulate social roleplay emerged from our interviews. Starting from the elder speakers, ANOVA revealed a radical opposition in VOT lengths between the subclasses M III NG and F III NG. What we call here F III NG is a social category often cited in perceptual dialectology<sup>36</sup> to typically hold a grudge against dialect and substandard linguistic traits, considered potentially humiliating and detrimental to the social promotion of the family. Our interviews delineated a similar picture: F III NG 2 admits being embarrassed by her inability to speak in St. Italian in shops and other everyday situations; F III NG 1 does not understand our interest in dialect and wishes it could just disappear. Directly connected is the scarce occurrence of the dialectal art. m. sg. def. form [i] in the speech of these two speakers: [i] is one of the few perceived trigger of RF, commonly evaluated as local and informal (Agostiniani, 1992). Our data thus suggest that M III NG > F III NG could be a simple case of sex-prestige pattern, characterized by the female avoidance of long-lag VOTs recognized as manly in the sense of substandard<sup>37</sup> and rustic. The absence of further specified ideological reinterpretation of the trait cannot be extended to the opposition observed in I, M I G > F I G. Our interviewed young Florentines showed a deep knowledge about the distribution and plausible contextual use of the trait, i.e. fit for informal conversation in a group of male-only friends. We can consider this kind of competence an example of what Iannàccaro (2002) calls *consapevolezza linguistica* (“linguistic awareness”), as it is opposed to *coscienza linguistica* (“linguistic consciousness”). Linguistic awareness about a trait implies a certain degree of metalinguistic knowledge about it, whereas linguistic consciousness does not: this is the case of the propensity shown by elder speakers to evaluate the trait during the matched-guise session, but not during the open-ended interviews<sup>38</sup>. It seems that

<sup>36</sup> For further references, see Iannàccaro (2002: 116-117).

<sup>37</sup> This claim is reinforced by the data reported by Soriano (1996: 137). VOT means of voiceless plosives in strong positions uttered by Florentine speakers of St. Italian are far shorter ([p:]: 9,25 ms.; [t:]: 14,25 ms.; [k:]: 27 ms.; [rp]: 10,5 ms.; [rt]: 15,50 ms.; [rk]: 26,25 ms.) than those of our Florentine dialect.

<sup>38</sup> It should be noted that we are interpreting here both *consapevolezza* and *coscienza linguistica* as two different levels of the same category of linguistic markers. The *gorgia enfatica* was never explicitly mentioned in the higher, more general and less induced layers of the open-ended interview; this leads us to think that the trait is not commonly available for overt commentary, not even for the young Florentine speakers (i.e. it is not a stereotype in the Labovian acceptance of the term).

young male Florentines intentionally use long-lag VOTs to index the belonging to a group of friends; in other words, the *gorgia enfatica* could represent an element of male *youth language* in Florence<sup>39</sup>. The use of substandard traits in male *youth language* to convey a sense of aggressiveness, rusticity and idealized manliness is a solid acquisition of dialectology (see, e.g. Cortelazzo, 1995: 585). The enthusiastic comments that we retrieved about the *gorgia enfatica* are a good proof of the *covert prestige* (Trudgill, 1972; Labov, 1990) that the trait holds among young men. From this point of view, the firm rejection of long lag VOTs by F I G speakers is not surprising. F I G is a subcategory of speakers typically associated with *overt prestige* traits (Cravens, Giannelli, 1995), considered the key of self-promotion and social acceptance. Nielsen (2011), developing the idea of finding traces of episodic memory in speech production, proved that VOT lengths can be a target of the so-called *imitation principle*: consistent exposure to lengthened VOTs induces the speakers to replicate the subsegmental trait in the production of new lexical items involving new types of plosives. In this sense, young people can easily take long-lag VOTs from a sensible distribution among the elders, and build an identity feature on them. The step from the evaluation of the trait among the elders to that of the young, idealizing manliness in the search for self-identity, is describable in terms of indexicality (Silverstein, 2003) by an opposition between a N<sup>th</sup> and a N<sup>th</sup> + 1 order<sup>40</sup>. The second age class remains with an indefinite status, conditioned by the personal biography of the single speaker and by the typology of *social networks* (Milroy, Milroy, 1978) linked to him. However, in what seems to be an *age-grading*<sup>41</sup> distribution hinting to a linguistic evolution during the lifespan of the speaker without stable age classes in the corresponding society, our third radical opposition, M II NG > M II G, could be interpreted as the snapshot of a *retrenchment* phenomenon, «a retreat from the non-standard variants used in youth by stabilization» (Evans Wagner, 2012: 375)<sup>42</sup>. We can conclude this section noticing how the potentiality of aspiration to index subsets of the primary less-marked ideological reference<sup>43</sup> could be considered an example of what Irvine, Gal (2000: 38) calls *fractal recursivity*, “the projection of an opposition, salient at some level of relationship, onto some other level”.

<sup>39</sup> Interestingly enough, Giacomelli (1934) describes a spontaneous conversation in a group of male young friends from Florence and Vicchio in which he heard frequent productions of voiceless aspirates.

<sup>40</sup> Even if the relation between the two orders is clear, we are not able to determine the exact indexical orders without further analysis about the linguistic ideologies associated to the trait. For some suggestions, see below.

<sup>41</sup> «Apparent time studies in which men are the most frequent users of a variant might be more indicative of age grading than change» (Evans Wagner, 2012: 374).

<sup>42</sup> Of course, the only way to confirm linguistic retrenchments would require a large-scale real-time research observing the behavior of a group of individuals during the passage from adolescence to adulthood (Chambers, 2003: 199). Our terminology should be intended as resulting from the interpretation of the distributional results of a cross-sectional study.

<sup>43</sup> In our case, the general idea of virility can further index the belonging to social subgroups of men not involved in “sophisticated” activities, and possibly the identification of boys in adolescent peer groups.

#### 4.2 Exaggeration as a pathway to manliness

Eckert, McConnell-Ginet (2003) gives great emphasis to the importance of exaggeration of biological differences in the construction of both masculine and feminine genders (see, e.g., op. cit.: 10, 13, 20). In particular, social psychology identifies a fundamental role of exaggeration in the expression of manliness. Desirable and acceptable masculine traits have a narrower definition than their feminine counterparts; moreover, masculinity tends to be more respected than femininity<sup>44</sup>. These facts lead manliness to an overexposure to social threats, and ultimately to defense mechanisms based on the so-called *masculine overcompensation thesis*. We saw how the use of dialectal traits is considered by young male speakers a viable metalinguistic stratagem to convey a sense of rudeness and idealized virility. However, in Florence the most marked local trait, the spirantizing *gorgia*, is far from being stigmatized. Recent studies (Marotta, 2014: 159; Biliotti, Calamai, 2012) suggest that a tendency to consider the *gorgia* an *overtly prestigious* trait is spreading throughout Tuscany. In this linguistic context, male *youth language* clearly needs to select, or create, something else. Italian linguistics reports examples of two different types of what we may call *linguistic exaggeration*. Franceschi (1969) analyses the historical consequences of *interdialectal* exaggeration: an observable overuse of a linguistic feature, targeting historically unmotivated contexts, could point to the late arrival of the feature in the diasystem as an erroneously replicated linguistic trend. In Franceschi's examples, we also find an ironical exaggerated imitation of the stereotypical *gorgia* by foreigners, [tos'k<sup>h</sup>ana] ("Tuscany"; Franceschi, 1969: 53). This seems to reflect the comments often generated by our unrealistically lengthened VOTs used in the matched-guise phase, hinting to the possibility of a folk reinterpretation of spirantized allophones in weak positions as aspirated plosives in strong positions. The other typology is *intradialectal* exaggeration. In 1979, Craffonara noticed some historically unexplainable palatalized realization of /ka, ga/ in Dolomitic Ladin. The author argues that those individual and irregular occurrences could be an attempt to convey a sense of local identity through overextension of the most recognizable dialectal feature, i.e. palatalization (Craffonara, 1979: 71). Even more pregnant for our exegetical attempt, Del Puente (1995) shows a clear decline of the metaphony in the dialect of Naples. However, few social categories prove the existence of a countertrend that even lead to an overextension of the metaphonized variants to etymologically unmotivated contexts. These categories consist in young low educated male speakers of the peripheral neighborhood and old male speakers of the same extraction. The trait holds *covert prestige* and is used to convey a sense of being an authentic Neapolitan man. These exaggerations are explicitly recognized and commented by the female counterparts; moreover, a *retrenchment* phenomenon is observed (Del Puente, 1995: 55-56), completing a picture strikingly similar to ours. Considering the formal implications of the example found in Franceschi (1969),

<sup>44</sup> See Willer, Rogalin, Conlon & Wojnowicz (2013: 984-985); Cheryan, Schwartz Cameron, Katagiri & Monin (2015).



and the referential ambivalence of the word *aspirazione*, that hints to a unified folk view of all the types of plosive allophones, *gorgia enfatica* could be born as a *covertly prestigious intradialectal* exaggeration of the overtly prestigious spirantization, the *gorgia*. In addition to that, aspiration seems to hold some ideological implications that make the trait an optimal candidate for the described purposes.

#### 4.3 Why aspiration? Plausible *addenda* to the indexical field

In 2012, Penelope Eckert formalized the birth of a third wave of studies on social meanings attributed to linguistic variation. The third wave does not link evoked social meanings to general sociolinguistic primary categories (first wave), nor to local categories (second wave), but believes that

variables cannot be consensual markers of fixed meanings; on the contrary, their central property must be indexical mutability. This mutability is achieved in stylistic practice, as speakers make social-semiotic moves, reinterpreting variables and combining and recombining them in a continual process of bricolage [...] (Eckert, 2012: 94).

This multilinear view of indexicality finds its main tool of formal representation in the *indexical field* model (Eckert, 2008), “a constellation of ideologically linked meanings, any region of which can be invoked in context” (Eckert, 2012: 94). Fully released, aspirated /t/ allophones are the core of one of the finest examples of the exegetical power of the *indexical field* (see, e.g., Eckert, 2008: 469; Drummond, Schlee, 2016: 57). These allophones in fact have been described in a good number of different social contexts as peculiar indexes of social meaning. Bucholtz (1996: 125) finds /t/ releases indexing intelligence and resistance to trendy speech styles in a group of high school “nerd” girls (and boys) in San Francisco. Benor (2001) attributes to the high frequency of aspirated final /t/ in the speech of boys in an Orthodox Jewish school of North Carolina a desired trait of erudition: aspiration is in fact not only distributed by gender, but also by level of instruction, being more frequent in students of the Yeshiva. In Podesva’s studies on gay speech (see, e.g., Podesva, 2007), the author observes the linguistic behavior of a gay doctor. This doctor produced more aspirated /t/s in the clinic than in informal situation, indexing precision and professionalism. However, during a barbecue with friends, the doctor uttered less fully released /t/s, but with much longer VOTs: he was using a parodic exaggeration of his professional persona, indexing exasperation fit to represent a “gay diva”. Eckert (2012: 97) expands on these findings suggesting that aspiration, as a fortition phenomenon, «can index emphasis or force, hence focus, power, or even anger». We found a similar phenomenon of *iconization* (Irvine, Gal, 2000) in some of our interviews. In an attempt to disambiguate the two differently perceived types of *aspirazione*, F II NG 2 defines [t<sup>h</sup>] as “aspirated... in the *strong* sense”; more significantly, evaluating the *gorgia enfatica*, M II G 2 says that “it’s a manly pronunciation... women prefer *weakened* forms”. Drummond, Schlee (2016) adds a proof to this argument with the indexical field reconstructed for the intervocalic t-glottalling, that shows a central branch tagged as “more laidback”. The authors

also try to unify the distributions of aspiration reported in the quoted studies noticing that “the meaning of the feature must relate to something shared by all these [groups or identities]; they all exploit indexical values linked to hyperarticulation (i.e. extremely articulate and clear speech)” (Drummond, Schlee, 2016: 53). We saw how hyperarticulation is often interpreted as a path to express intelligence and precision: from this point of view, the distribution and evaluation of the *gorgia enfatica* as rustic and informal seems at least controversial. However, as Podesva (2013: 428) argues about non-modal phonation types, “social meanings [...] are culturally specific and should not be reduced to purely iconic or unanalyzed associations [...]”. Hyperarticulation is also a correlate of *stancetaking* (Freeman, 2014), intended here as “modality, evaluation, subjectivity, epistemicity, footing, alignment, assessment, agreement, and so on, [used] to refer to a speaker’s or writer’s attitude, displays of emotions and desires, expressions of beliefs and certainty toward given issues, people, and the speakers’ coparticipants” (Haddington, 2004: 103). The comments of our speakers suggest some level of *stance accretion*<sup>45</sup> based on this correlation. M III G 2 thinks that the *gorgia enfatica* is a virile communicative strategy to stress an opinion, enriched by a feeling of irritation. M I N G 2 is surprised by the perceived coincidence between a trait used to clarify concepts in professional environments and an acoustic memory of aspirations used in friendly speech during the adolescence. M II G 2 is confused by a trait that conveys at the same time dialectality and firmness, precision. We can thus assume that Florentine men may find in long-lag VOTs a way to express an identity of firm stance-takers, decision-makers, ideologically dynamic people.

## 5. Conclusions

Shifting from production to perception, from sex to gender, we noticed how an instrumentally retrieved, statistically significant distribution of an allophone could potentially be the core of an ideological constellation, imbued with social meanings at various levels of analysis. From this point of view, Oh’s suggestion to link distributional studies on VOT-sex patterns to in-depth sociophonetic descriptions of the chosen speech community is still actual and exegetically promising. Much work is yet to do on both the production and the perception of the *gorgia enfatica*; however, the integration of these two aspects of the research in a unified protocol proved itself useful for a first level of description of the social factors conditioning the observable distribution of aspirated voiceless plosives in the Tuscan chief-town. “People (perhaps especially men) perform gender” (Johnson, 2006: 487), and sociophonetics has the potential to explain these performances.

<sup>45</sup> «The way in which stances accumulate in more durable structures of identity» (Bucholtz, Hall, 2005: 596). In other words, it is for example the process through which people frequently showing joyful behaviors can be seen *tout court* as joyful.

### *Bibliography*

- ABUDALBUH, M. (2011). Effects of gender on the production of emphasis in Jordanian Arabic: A sociophonetic study. In *Kansas Working Papers in Linguistics*, 32, 20-47.
- AGOSTINIANI, L. (1992). Su alcuni aspetti del "rafforzamento sintattico" in Toscana e sulla loro importanza per la qualificazione del fenomeno in generale. In *Quaderni di Linguistica dell'Università di Firenze*, 3, 1-28.
- ALZOUBI, A. (2016). A sociophonetic study of the effect of gender on emphatic-plain contrast in Jordanian Arabic. In *Journal of the Acoustical Society of America*, 140, 3113.
- BELL, A. (1984). Language style as audience design. In *Language in Society*, 13, 145-204.
- BENOR, S. (2001). Sounding learned: the gendered use of /t/ in Orthodox Jewish English. In JOHNSON, D.E., SANCHEZ, T. (Eds.), *Penn working papers in linguistics: selected papers from NWA 29*. Philadelphia: University of Pennsylvania, 1-16.
- BIJANKHAN, M., NOURBAKHS, M. (2009). Voice onset time in Persian initial and intervocalic stop production. In *Journal of the International Phonetic Association*, 39(3), 335-364.
- BILIOTTI, F., CALAMAI, S. (2012). Linguistic opinions and attitudes in Tuscany: verbal guise experiments on the varieties of Arezzo and Florence. In CALAMAI, S., CELATA, C. & CIUCCI, L. (Eds.), *Sociophonetics, at the crossroads of speech variation, processing and communication*. Pisa: Edizioni della Scuola Normale Superiore, 1-4.
- BRINCA, L., ARAÚJO, L., NOGUEIRA P. & GIL, C. (2016). Voice onset time characteristics of voiceless stops produced by children with European Portuguese as mother tongue. In *Ampersand*, 3, 137-142.
- BUCHOLTZ, M. (1996). Geek the girl: language, femininity and female nerds. In WARNER, N., AHLERS, J., BILMES, L., OLIVER, M., WERTHEIM, S. & CHEN, M. (Eds.), *Gender and belief systems*. Berkeley: Berkeley Women and Language Group, 119-131.
- BUCHOLTZ, M., HALL, K. (2005). Identity and interaction: a socio-cultural linguistic approach. In *Discourse Studies*, 7, 585-614.
- CAMPBELL-KIBLER, K. (2007). Accent, (ING), and the social logic of listener perceptions. In *American Speech*, 82, 32-64.
- CALAMAI, S. (2015). *Introduzione alla sociofonetica*. Roma: Carocci.
- CALAMAI, S., RICCI, I. (2005). Un esperimento di matched-guise in Toscana. In *Studi Linguistici e Filologici on Line*, 3(1), 63-105.
- CASTELLANI, A. (1952). *Nuovi testi fiorentini del Dugento*. Firenze: Sansoni.
- CASTELLANI, A. (1960). Precisazioni sulla gorgia toscana. In *Boletim de Filologia*, XIX, 242-262.
- CHAMBERS, J.K. (2003), *Sociolinguistic Theory*. 2nd edition. Oxford: Blackwell.
- CHENG, M. (2013). Voice Onset Time of Syllable-Initial Stops in Sixian Hakka: Isolated Syllables. In *Journal of National Taiwan Normal University*, 58(2), 193-227.
- CHERYAN, S., SCHWARTZ CAMERON, J., KATAGIRI, Z. & MONIN, B. (2015). Manning Up. Threatened Men Compensate by Disavowing Feminine Preferences and Embracing Masculine Attributes. In *Social Psychology*, 46, 218-227.

- COLLINS, K.M.T., ONWUEGBUZIE, A.J. & SUTTON, I.L. (2006). A Model Incorporating the Rationale and Purpose for Conducting Mixed-Methods Research in Special Education and Beyond. In *Learning Disabilities: A Contemporary Journal*, 4(1), 67-100.
- CORTELAZZO, M.A. (1995). La componente dialettale nella lingua delle giovani e dei giovani. In MARCATO, G. (Ed.), *Donna & linguaggio*. Padova: CLEUP, 581-586.
- CRAFFONARA, L. (1979). Zur Palatalisierung von CA und GA in den Sellatälern. In *Ladinia*, III, 69-93.
- CRAVENS, T.D., GIANNELLI, L. (1995). Relative salience of gender and class in a situation of multiple competing norms. In *Language Variation and Change*, 7, 261-285.
- DEL PUENTE, P. (1995). La metafonía napoletana: Un tentativo di analisi sociolinguistica. In *L'Italia dialettale*, 58, 49-68.
- DRUMMOND, R., SCHLEEF, E. (2016). Identity in variationist sociolinguistics. In PREECE, S. (Ed.), *The Routledge Handbook of Language and Identity*. London: Routledge, 50-65.
- ECKERT, P. (2008). Variation and the indexical field. In *Journal of Sociolinguistics*, 12(4), 453-476.
- ECKERT, P. (2012). Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. In *Annual Review of Anthropology*, 41, 87-100.
- ECKERT, P., MCCONNELL-GINET, S. (2003). *Language and Gender*. Cambridge: Cambridge University Press.
- EVANS WAGNER, S. (2012). Age Grading in Sociolinguistic Theory. In *Language and Linguistic Compass*, 6(6), 371-382.
- FRANCESCHI, T. (1969). Il principio dell'esagerazione come criterio di ricerca linguistica. In *Archivio Glottologico Italiano*, LIV, 49-105.
- FRANCIS, A.L., CIOCCA, V. & YU, J.M.C. (2003). Accuracy and variability of acoustic measures of voicing onset. In *Journal of the Acoustical Society of America*, 113(2), 1025-1032.
- FREEMAN, V. (2014). Hyperarticulation as a signal of stance. In *Journal of Phonetics*, 45, 1-11.
- GIACOMELLI, R. (1934). Controllo fonetico per diciassette punti dell' AIS nell' Emilia, nelle Marche, in Toscana, nell' Umbria e nel Lazio. In *Archivum Romanicum*, XVIII, 153-211.
- GIANNELLI, L., (1976). *Toscana*. Pisa: Pacini.
- GIANNELLI, L., (1983). Aspirate etrusche e gorgia toscana: valenza delle condizioni fonetiche dell' area toscana. In AGOSTINIANI, L., GIANNELLI, L. (Eds.), *Fonologia etrusca, fonetica toscana: il problema del sostrato*. Firenze: Olschki, 61-102.
- GIANNELLI, L., SAVOIA, L.M. (1978). L' indebolimento consonantico in Toscana. In *Rivista Italiana di Dialettologia*, II, 23-58.
- GIANNELLI, L., SAVOIA, L.M. (1979-1980). L' indebolimento consonantico in Toscana. In *Rivista Italiana di Dialettologia*, IV, 38-101.
- HADDINGTON, P. (2004). Stance Taking in News Interviews. In *SKY Journal of Linguistics*, 17, 101-142.
- HAN, K. (2010). The Interaction between Phonetic Environment and Speaker Sex on VOT Differences. In *The Journal of Linguistic Science*, 54, 295-308.

- HUDSON, R.A. (1996). *Sociolinguistics*. 2nd edition. Cambridge: Cambridge University Press.
- HUGGINS, A.W.F. (1972). Just Noticeable Differences for Segment Duration in Natural Speech. In *The Journal of the Acoustical Society of America*, 51(4), 1270-1278.
- IANNACCARO, G. (2002). *Il Dialetto Percepito. Sulla reazione di parlanti di fronte al cambio linguistico*. Alessandria: Edizioni dell'Orso.
- IRVINE, J.T., GAL, S. (2000). Language Ideology and Linguistic Differentiation. In KROSKRITY, P.V. (Ed.), *Regimes of Language*. Santa Fe, NM: Sch. Am. Res. Press, 35-83.
- IZZO, H.J. (1972). *Tuscan & Etruscan*. Toronto: University of Toronto Press.
- JOHNSON, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. In *Journal of Phonetics*, 34, 485-499.
- KANG, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. In *Journal of Phonetics*, 45, 76-90.
- KIESLING, S.F. (2013). Constructing identity. In CHAMBERS, J.K., SCHILLING, N. (Eds.), *The handbook of language variation and change*. 2nd ed. Malden: John Wiley and Sons, 448-467.
- KOENIG, L.L. (2000). Laryngeal Factors in Voiceless Consonant Production in Men, Women, and 5-Year-Olds. In *Journal of Speech, Language, and Hearing Research*, 43, 1211-1228.
- KRISTIANSEN, T. (2010). Investigating language in space: Experimental techniques. In AUER, P., SCHMIDT, J.E. (Eds.), *Language and Space: An International Handbook of Linguistic Variation. Volume 1: Theories and Methods*. Berlin: Walter de Gruyter, 528-549.
- LABOV, W. (1984). Field methods of the Project on Linguistic Change and Variation. In BAUGH, J., SHERZER, J. (Eds.), *Language in Use*. Englewood Cliffs: Prentice-Hall, 28-53.
- LABOV, W. (1990). The intersection of sex and social class in the course of linguistic change. In *Language Variation and Change*, 2, 205-251.
- LABOV, W. (1994). *Principles of Linguistic Change. Volume 1: Internal Factors*. Oxford: Blackwell.
- LAMBERT, W.E. (1967). The social psychology of bilingualism. In *Journal of Social Issues*, 23, 91-109.
- LEE, H. (2016). Consistent sound change between stops and affricates in Seoul Korean within and across individuals: A diachronic investigation. In *Journal of the Acoustical Society of America*, 140, 491-496.
- LI, F. (2013). The effect of speakers' sex on voice onset time in Mandarin stops. In *The Journal of the Acoustical Society of America*, 133, 142-147.
- LISKER, L., ABRAMSON, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. In *Word*, 20(3), 527-565.
- LOPORCARO, M. (1997). Lengthening and raddoppiamento fonosintattico. In MAIDEN, M., PARRY, M. (Eds.), *The Dialects of Italy*. London and New York: Routledge, 41-51.
- LUNDEBORG HAMMARSTRÖM, I., LARSSON, M., WIMAN, S. & McALLISTER, A. (2012). Voice onset time in Swedish children and adults. In *Logopedics, Phoniatrics, Vocology*, 37(3), 117-122.
- MADHU, S.R.B., MAHENDRA, K.N. & SREEDEVI, N. (2014). Voice Onset Time across Gender and Different Vowel Contexts in Telugu. In *Language in India*, 14(12), 252-263.

- MARCATO, G., URSINI, F. & POLITI, A. (1974). *Dialetto e italiano: status socioeconomico e percezione sociale del fenomeno linguistico*. Pisa: Pacini.
- MAROTTA, G. (2004). Non solo spiranti. La gorgia toscana nel parlato di Pisa. In *L'Italia Dialettale*, 62, 27-60.
- MAROTTA, G. (2014). New parameters for the sociophonetic indexes. Evidence from the Tuscan varieties of Italian. In CELATA, C., CALAMAI, S. (Eds.), *Advances in Sociophonetics*. Amsterdam-Philadelphia: John Benjamins, 137-168.
- MILROY, J., MILROY, L. (1978). Belfast: Change and Variation in an Urban Vernacular. In TRUDGILL, P. (Ed.), *Sociolinguistic Patterns in British English*. Baltimore: University Park Press, 19-36.
- MILROY, L. (1987). *Language and Social Networks*. Oxford: Blackwell.
- MIONI, A.M. (2001). *Elementi di fonetica*. Padova: Unipress.
- MOLFESE, D.L., HESS, T.M. (1978). Hemispheric Specialization for VOT Perception in the Preschool Child. In *Journal of Experimental Child Psychology*, 26, 71-84.
- MOLFESE, D.L., MOLFESE, V.J. (1988). Right hemisphere responses from preschool children to temporal cue to speech and nonspeech materials: electrophysiological correlates. In *Brain and Language*, 33, 245-259.
- MORRIS, R.J., MCCREA, C.R. & HERRING, K.D. (2008). Voice onset time differences between adult males and females: Isolated syllables. In *Journal of Phonetics*, 36, 308-317.
- NAKAI, S., SCOBIE, J.M. (2016). The VOT category boundary in word-initial stops: Counter-evidence against rate normalization in English spontaneous speech. In *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1), 1-31.
- NIELSEN, K. (2011). Specificity and abstractness of VOT imitation. In *Journal of Phonetics*, 39, 132-142.
- NISSEN, S.L., FOX, R.A. (2009). Acoustic and spectral patterns in young children's stop consonant productions. In *Journal of the Acoustical Society of America*, 126, 1369-1378.
- NODARI, R. (2015). Descrizione acustica delle occlusive sorde aspirate: analisi sociofonetica dell'italiano regionale di adolescenti calabresi. In VAYRA, M., AVESANI, C. & TAMBURINI, F. (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio. Studi AISV 1*. Milano: Officinaventuno, 139-153.
- OH, E. (2011). Effects of speaker gender on voice onset time in Korean stops. In *Journal of Phonetics*, 39, 59-67.
- PENG, J., CHEN, L. & LEE, C., (2014). Voice Onset Time of Initial Stops in Mandarin and Hakka: Effect of Gender. In *Taiwan Journal of Linguistics*, 21(1), 63-80.
- PICCARDI, D. (in press). La cosiddetta "gorgia enfatica" fiorentina: una ridefinizione in chiave sociofonetica. In *L'Italia Dialettale*, 79.
- PICKETT, J.M. (1980). *The sounds of speech communication: A Primer of Acoustic Phonetics and Speech Perception*. Austin: University Park Press.
- PODESVA, R. (2007). Phonation type as a stylistic variable: the use of falsetto in constructing a person. In *Journal of Sociolinguistics*, 11, 478-504.
- PODESVA, R. (2013). Gender and the social meaning of non-modal phonation types. In CATHCART, C., CHEN, I., FINLEY, G., KANG, S., SANDY, C.S. & STICKLES, E. (Eds.), *Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society*, 427-448.

- ROBB, M., GILBERT, H. & LERMAN, J. (2005). Influence of gender and environmental setting on voice onset time. In *Folia Phoniatrica et Logopaedica*, 57(3), 125-133.
- ROMITO, L., CIARDULLO, M.A. & TARASI, A. (2015). Analisi acustica delle occlusive sorde aspirate del dialetto di San Giovanni in Fiore (CS). In VAYRA, M., AVESANI, C. & TAMBURINI, F. (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio. Studi AISV I*. Milano: Officinaventuno, 169-186.
- RYALLS, J., ZIPPRER, A. & BALDAUFF, P. (1997). A preliminary investigation of the effects of gender and race on Voice Onset Time. In *Journal of Speech and Hearing Research*, 40(3), 642-645.
- SCHARF, G., MASUR, H. (2002). Voice onset time in normal speakers of a German dialect. Effects of age, gender and verbal material. In WINDSOR, F., KELLY, M.L. & HEWLETT, N. (Eds.), *Investigations in Clinical Phonetics and Linguistics*. Mahwah: Lawrence Erlbaum, 327-339.
- SILVERSTEIN, M., (2003). Indexical order and the dialectics of sociolinguistic life. In *Language & Communication*, 23, 193-229.
- SINGH, R., KESHET, J., GENCAGA, D. & RAJ, B. (2016). The relationship of voice onset time and voice offset time to physical age. In *Proceedings of the 41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016)*, Shanghai, China, 5390-5394.
- SKOOG WALLER, S., ERIKSSON, M. & SÖRQVIST, P. (2015). Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. In *Frontiers in Psychology*, 6, 978.
- SMITH, B.L. (1978). Effects of Place of Articulation and Vowel Environment on "Voiced" Stop Consonant Production. In *Glossa*, 12(2), 163-175.
- SORIANELLO, P. (1996). Indici fonetici delle occlusive sorde in cosentino. In *Rivista Italiana di Dialettologia*, XX, 123-159.
- SOUKUP, B. (2013). On matching speaker (dis)guises – revisiting a methodological tradition. In KRISTIENSEN, T., GRONDELAERS, S. (Eds.), *Language (de)standardisation in late modern Europe: Experimental studies*. Oslo: Novus, 267-285.
- STEVENS, K.N. (1999). *Acoustic Phonetics*. Cambridge: MIT Press.
- SUMMERFIELD, A.Q. (1975). Aerodynamics versus mechanics in the control of voicing onset in consonant-vowel syllables. In *Speech Perception*, Series 2, 4, 61-72.
- SWARTZ, B.L. (1992). Gender difference in voice onset time. In *Perceptual and Motor Skills*, 75, 983-992.
- SWEETING, P.M., BAKEN, R.J. (1982). Voice onset time in a normal-aged population. In *Journal of Speech and Hearing Research*, 25, 129-134.
- TITZE, I.R. (1994). *Principles of Voice Production*. Englewood Cliffs: Prentice Hall.
- TRUDGILL, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. In *Language in Society*, 1(2), 179-195.
- WHITESIDE, S.P., HENRY, L. & DOBBIN, R. (2004). Sex differences in voice onset time: a developmental study of phonetic context effects in British English. In *Journal of the Acoustical Society of America*, 116(2), 1179-1183.

- WHITESIDE, S.P., IRVING, C.J. (1997). Speakers' sex differences in voice onset time: Some preliminary findings. In *Perceptual and Motor Skills*, 85(2), 459-463.
- WHITESIDE, S.P., IRVING, C.J. (1998). Speaker's sex differences in voice onset time: A study of isolated word production. In *Perceptual and Motor Skills*, 86, 651-654.
- WHITESIDE, S.P., MARSHALL, J. (2001). Developmental Trends in Voice Onset Time: Some Evidence for Sex Differences. In *Phonetica*, 58, 196-210.
- WILLER, R., ROGALIN, C.L., CONLON, B. & WOJNOWICZ, M.T. (2013). Overdoing Gender: A Test of the Masculine Overcompensation Thesis. In *American Journal of Sociology*, 118(4), 980-1022.
- YU, V.Y., DE NIL, L.F. & PANG, E.W. (2015). Effects of age, sex and syllable structure on voice onset time: Evidence from children's voiceless aspirated stops. In *Language and speech*, 58(2), 152-167.
- YUAN, J., LIBERMAN, M. & CIERI, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *Proceedings of the 9th International Conference on Spoken Language Processing (Interspeech-ICSLP)*, 541-544.