'A SCOUSE VOICE? HARSH AND UNFRIENDLY!'

PHONETIC CLUES TO THE PERCEPTION OF VOICE QUALITY IN LIVERPOOL ENGLISH

MASSIMILIANO BARBERA & MARLEN BARTH *

1. THE SCOUSE

1.1 Historical and linguistic aspects

The term Scouse indicates the variety of English spoken in Liverpool and Merseyside. This accent, which is immediately recognized by native speakers of English, presents various typically Irish English\(^1\) features (cf. § 2.2), whereas it is noticeably different from the accents of the surrounding areas. Besides that, there is a number of other facts that make a possible Irish origin of the Scouse accent plausible\(^2\). First of all, the geographical position of Liverpool on the north-west coast of England has fostered the contact between the city on the Mersey and Ireland since the Middle Ages. The rise of modern Scouse, however, corresponds to the period of the so-called Irish potato famine, which in the 1840s forced thousands of Irish to leave their country. This migration invested Liverpool in the first place and can be verified in the demographic data of those years. In 1841, for instance, 17.3% of all immigrants (44.9%) were born in Ireland. Although there were also migrations from Wales and Scotland, the Irish was undoubtedly the most important one and

\(^1\) This term indicates the variety English spoken in Ireland.

\(^2\) For a detailed discussion of the demographic and historical data of the Irish immigration see Marotta (2006).
continued for the entire 20th century. Therefore, we can suppose that still today many inhabitants of Liverpool are of Irish origin.

Due to their limited financial resources, the Irish immigrants went to live in the poor parts of the city, mainly near the port. Apart from this socio-economic segregation, there were also religious differences (English Protestants vs Irish Catholics) as well as differences in the language, since the variety of English spoken by the immigrants was strongly influenced by the Irish interference and hence contrasted with north-western English, which constituted the standard local variety.

It is therefore easily understandable that for the Irish immigrants we can speak of a close-knit social network (cf. Milroy, 1980, 2002), for a long time without contact with other social networks. Thus, the characteristic features of the Anglo-Irish variety could be preserved. In this regard it is important to underline how the combination of socio-economic and cultural facts has determined the discrimination of the Irish immigrants and of the variety of English spoken by them.

It was only in the middle of the 20th century, mainly during the Second World War, that growing interactions between the autochthonous community and the immigrants lead to a loosening of the close-knit social network. As a result of the more frequent contacts between the two communities, some features of the language of the Irish, stigmatized up to then, entered the standard local variety becoming endemic. One of the factors which contributed to increase the prestige of Scouse in Great Britain was certainly the success of the

---

3 In this regard, the move of the BBC studios to Liverpool during the war has certainly played an important role.
band *The Beatles* in the Sixties. Moreover, recently, Liverpool has turned into Britain’s favourite location for call centres (cf. Ward, 2000). Even though it remains partly stigmatized, a Scouse voice is no longer perceived as harsh, but as straight, understanding and friendly as well as having covert prestige.

1.2 Segmental and suprasegmental features

Although Scouse is marked mostly at the phonetic and prosodic levels, it also peripherally involves the lexical and morphosyntactic levels. The Scouse accent affects both the vowel system and the consonant system.

As far as the vowel system is concerned, there is a series of neutralizations of phonemic oppositions present in RP. A typical Scouse feature is the pronunciation of the central open-mid vowel /ɜː:/ as [ɛː] or as diphthong [ɛə], which leads to homophony between words like *her* and *hair*. Also the realization of the close front vowel /i/ as long and tense in all contexts is characteristic of the Liverpool accent. The pronunciation of the open-mid back vowel /ʌ/ as [u], on the other hand, is typical for all northern varieties of English. The outcome of this phenomenon is the neutralization of minimal pairs in RP, such as for instance *luck* and *look*. Also /æ/ is altered in Scouse, becoming a back vowel [a].

As regards the consonant system, the presence of typical Irish English features seems to indicate a possible Irish origin of Scouse. In
this regard, one of the characteristics that *Scouse* shares with Irish English is the pronunciation of the interdental fricatives /θ/ and /ð/ as dental stops [t] and [d]. Furthermore, it is probable that also the realization of the velar nasal [ɳ] as cluster [ŋŋ] in word-final context is of Irish English origin.

### 1.2.1 Lenition

The feature which more than any other characterizes the *Scouse* consonant system is the lenition of the stop segments. The same phenomenon is present in Irish English too. Initially being described as typical of the working class, the consonantal weakening seems to spread out, affecting also the speech of higher classes (cf. Marotta, 2004). Moreover, lenition seems to be spreading to formerly not affected phonemes, such as the voiced alveolar stop

4. Lenition in *Scouse* is of the opening type, i.e. no change in voicing occurs. The weakening is rather due to a decreasing resistance to the air-flow in the vocal tract5. The outcome of lenition of stop phonemes in *Scouse* are affricates or fricatives. According to Honeybone (2001) all obstruents in Liverpool English /p b t d k g/ can potentially be target of lenition, although some of them seem to undergo the process more frequently; the voiceless alveolar is the most affected, followed by the voiceless velar, while the voiceless bilabial and the voiced alveolar are lenited less frequently and only in special contexts.

---

Although there are differences between the single phonemes as far as favouring contexts are concerned, the weakening occurs most frequently in intervocalic position in unstressed syllable.

The segment which, apart from being the one most frequently lenited, also presents the highest number of allophones, ranging from the affricate to the fricative and the glottal approximant – permitting even elision in monosyllabic function words (such as but, what, it)\(^6\) – is the voiceless alveolar stop /t/. A possible explanation for this fact is the so-called coronal syndrome (cf. Kenstowicz, 1994: 516). An interesting allophone output of the lenition of /t/ is the fricative. This sound, called slit fricative by Pandeli et al. (1997), is produced without contact between the tongue and the alveolar ridge, since the tongue shape is flat cross-sectionally; in this way, a broad fricative channel is created in the vocal tract. The symbol used for this slit fricative is [ɭ], that is the symbol of the voiceless interdental fricative as base symbol together with the diacritic for a dental place of articulation taken from the extended IPA for disordered speech.

\(^6\) Cf. Watson (2002).
Figure 1. Waveform and spectrogram of the phrase She was very unhappy that she was not invited; subject DS (F); intervocalic /t/ realized as [ð].

Recently, it was argued that there are slit allophones also for the voiced alveolar stop. These sounds, transcribed as [ð] for the affricate and [θ] for the fricative, show the same acoustic features as [t̪] and [θ], since they differ from them only in that they are less intense and shorter and have the typical sonority bar. With reference to the transcription already in use for [t̪] and [θ], we propose an analogous graphic representation for the homorganic voiced sounds⁷.

⁷ For a more detailed discussion of this topic see Marotta & Barth (2006).
As regards the prosodic features, Scouse differs from RP for the rising tone at the end of declarative sentences, instead of a falling tone, like in RP. This Final Rising Pattern is common to the so-called Urban Northern British English (cf. Cruttenden 1994): the varieties of English spoken in Northern Ireland and Western Scotland, as well as in the cities of Birmingham, Newcastle and Liverpool do share this melodic pattern. Thus, in these varieties interrogative sentences differ from declarative ones only for the greater frequency range of the former. It has been hypothesized that this melodic pattern is due to the Celtic background which the above mentioned places have in common: all of them have been a target of immigration for a high number of Celtic populations during the 19th century.
1.3 The articulatory setting

The most relevant aspect of the Scouse accent is probably its phonatory and articulatory setting. The most detailed description of this special setting is certainly Knowles (1973). According to the author, in the Scouse articulation, the pharynx is tightened and the larynx is displaced upwards, while the lower jaw is held close to the upper jaw even in the pronunciation of open vowels. The tongue is raised and retracted, with the back raised to the velum, thus increasing the volume in the front part of the oral tract. At the same time this leads to a constriction in the back of the oral cavity which forces the velum in an intermediate position. Due to the immobility of the tongue caused by this setting, its tip intervenes also in the articulation of alveolar segments. So, it is rather obvious that the velo-pharyngeal mechanism is used in an unnatural way. According to Knowles, there
is a correlation between this setting and the velarization of consonants in Scouse.

Another feature that characterizes the Scouse articulatory setting is the relaxation of the lower lip and the tongue (lax voice), which impedes the complete closure of the organs during the articulation of stops.

1.4 Voice quality

A Scouse voice is often described as “adenoidal” or “nasal”, i.e. produced with a partial obstruction in the nasal tract, leading to denasalization. It is important to underline that this voice quality requires an unnatural and hence uneconomical use of the velo-pharyngeal mechanism. This is in contrast with the definition of the accent as “lax”. This seems to be the main reason why a Scouse voice is perceived as “ugly” and “unfriendly” by native speakers of English.

As regards the position of the velum during articulation, Cagliari (1978; in Laver 1980) suggests a neutral scale of velarity according to which the height of the velum varies according to the segment produced. Moreover, the author supposes that in the movement from a neutral velo-pharyngeal setting to a nasal one, at least some segments have to show a lowering of the velum compared to the values of the neutral scale, while a denasalized voice quality requires its raising.
2. EXPERIMENTAL ANALYSIS

2.1 The corpus

The present study was based on the analysis of six subjects of Liverpool English, judged as representative speakers of this accent. Two sociolinguistic variables guided our choice of the speakers: their age and their gender. Since the data set is relatively small, we decided to choose only adolescents of approximately the same age in order to permit comparison of the results without having to pay attention to many variables. Another social factor that seems to affect variation within dialects is the speaker’s gender. For this reason we decided to analyze the speech of both males and females.

The subjects, who at the moment of the recording were between sixteen and seventeen years old, were all born and have grown up in Liverpool. Three of them were male (GW, NS, PH) and three were female (LL, LM, DS). The recordings of five of the six speakers (GW,
NS, PH, LM, DS) were taken from the IViE corpus, whereas speaker LL was recorded in Viareggio (Lucca, Italy) in August 2004.

The subjects of the IViE corpus were asked to re-tell the fairytale of Cinderella, which they had previously read. The spontaneous speech for speaker LL was obtained in a conversation, asking her questions about informal topics.

2.2 Methodology

The acoustic analysis was carried out in the Laboratory of Phonetics of the Department of Linguistics, University of Pisa, using the software Praat (version 4.3.29).

For each speaker we analyzed six strings of spontaneous speech containing at most three breath groups. In these strings the following parameters were measured, first for the entire breath group and then for single sonorant segments and contiguous vowels:

a) the mean fundamental frequency;

b) the Jitter (local %), indicating the average absolute difference between consecutive periods, divided by the average period;

c) the Shimmer (local %), i.e. the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude;

d) the HNR (dB), the proportional degree of acoustic aperiodicity between the harmonics and the noise.

---

8 The abbreviation IViE stands for Intonational Variation in English, a project carried out by Esther Grabe, Brechtje Post and Francis Nolan in 2001. The IViE corpus is available on-line at the website www.phon.ox.ac.uk.
Nasal consonants are of course the favourite target of denasalization. So, we took them as a first point of reference. Since also the laterals presented spectro-acoustic characteristics very similar to those found for the nasals, we decided to analyze them too. The articulation of these sounds implicates a forward movement of the tongue and hence of the velum and has therefore a velo-pharyngeal and lingual setting comparable to that of the nasals. On the other hand, we omitted the vibrants, for they have a rather strong intrinsic noisy component which may forge the reliability of the above mentioned parameters. We analyzed the vowels since they are affected by coarticulatory effects induced by the contiguous consonantal segments; being phones theoretically lacking aperiodicity, they should show structural alterations connected with the phono-articulatory supra-laryngeal setting rather clearly.

3. THE PHONATORY SETTING

3.1 A peculiar structure

From the very beginning, the acoustic analysis revealed the presence of a peculiar structure with a morphology undoubtedly due to friction noise. The chosen segments (sonorants and vowels) are characterized nearly systematically by a special structure in the spectrogram: fine vertical lines arranged in parallel horizontally indicate the presence of an anomalous resonance (cf. Fig. 5).

If we keep in mind the comments made in the literature on the general Scouse phonatory setting (cf. §§ 2.3 and 2.4) and if we
combine them with our observations, it seems to be possible to identify a particular supra-laryngeal setting able to justify the acoustic effect (improperly) defined as nasal as well as the just described evidence from the spectrogram. We could schematize as follows:

a) the velum is in relatively lowered position, thus permitting a limited, but constant air-flow from the nasal cavities due to the only partial obstruction by the velum;

b) the signs in the spectrum (the “little bricks” clearly visible in Fig. 5) might hence be attributed to two connected factors: the friction in the velo-pharyngeal tract together with the vibration of the velum itself; this last point would give a kind of regularity to the aperiodic structure produced by the friction;

c) the nasal segments (but, for the above mentioned reasons, also the contiguous vowels and the lateral phones) are strongly affected by these effects: the velo-pharyngeal constriction seems to reduce the normal nasal resonance;

d) therefore, the diffused velo-pharyngeal friction can be identified even in the spectrogram.
This hypothesis is confirmed also by Knowles (1973: 12): *The efficiency of the nose as a resonator for the nasals depends on the degree of opening, and it is possible that it is habitually less open in Scouse than RP, and may be so close as to impede slightly the air-flow.*

The partial obstruction created by the velum in mid-lowered position together with its vibration – caused by the air-flow coming from the nasal cavities – would hence produce a particular vocal compound generally qualifiable as hypo-nasalized rather than denasalized. In other words, the overall perception is affected by this peculiar trait of nasality, which, however, indicates an anomalous use of the velo-pharyngeal tract as regards the phonatory dynamics and above all for the effects on the really nasal segments (cf. Fig. 7).
For this reason, the *Scouse* voice quality should be interpreted as a special colour, to the formation of which contribute synergically the oral as well as the nasal output (cf. Fig. 8).

![Figures 6 and 7. Anatomic localization and schematic representation of the velo-pharyngeal friction.](image)

3.2 Nasal or adenoidal?

The label *nasal voice* is basically linked to the acoustic percept (Laver, 1980: 86); in other words, this definition corresponds to the subjective recognition of nasal resonance as characteristic of the speaker’s overall colour of voice. In fact, from an articulatory point of view we can identify different velo-pharyngeal dynamics which justify a perceptive output qualifiable as *nasal voice*, *adenoidal* or, in medical terms, *rinolalia*. However, it should be underlined that a really nasal voice has a phonetic correlate of specific phonatory processes which actually cause an increment of the nasal resonance. On the spectro-acoustic level, this peculiarity is visible most of all in
the vowel segments, which, of course, are affected by the effect of articulatory contamination produced by a considerable air-flow through the nose. An adenoidal voice, on the other hand, is by definition denasalized and hence marked, as already mentioned, by perceptual features caused by a partial obstruction of the nasal cavities; on the whole, the colour is similar to that of speakers with adenoids. Despite this contrasting aetiology, the two mechanisms have been settled on the same definition based mainly on perception, namely nasalization.

If we consider these premises, the definition of the Scouse voice as either adenoidal or nasal appears inadequate or at least imprecise. First of all, the double denomination does not resolve the basic ambiguity of the actual dynamics underlying this particular voice quality. Moreover, the possibility to distinguish between two phonatory outputs, which are in fact similar, but however not identical, remains vague. As regards the Scouse accent, we should presume a reduction of the nasal resonance triggered by a special phono-articulatory mechanism, keeping in mind the energetic balance system of the vocal and verbal production as a whole.

In the light of the analysis carried out and above all because of the spectral morphology found in the subjects’ speech, we propose – as already mentioned – the label hypo-nasalized voice, which describes in a more adequate manner the actual anomaly (both on the perceptive and the spectro-acoustic level) of the vowels and the sonorant segments. For articulatory contiguity, these sounds activate the velo-
pharyngeal friction in the supra-laryngeal tract to a higher degree than a voice quality classifiable as neutral or modal.

3.3 A hypothesis: the scales of energy

At this point it is necessary to reconcile the connotation lax of the articulatory setting – confirmed by the large presence of consonantal weakening – with a phonatory setting undoubtedly definable as tense voice. If we take into account the phonopoietic act as a whole, we can suppose that the total energetic expenditure is subject to an intrinsic balancing according to which to a reduction of the use of energy at the articulatory level corresponds an increase of the use of energy at the laryngeal level (causing instances of creaky voice), but above all at the supra-laryngeal level (related to a hyperactivation of the velo-pharyngeal mechanism).

In fact, as regards the articulatory output, lenition indicates a reduced use of energy (corresponding to an overall relaxed setting of the external articulatory organs, while at the phono-articulatory level the velo-pharyngeal tract and the tongue take an unnatural setting which implicates a higher effort.

Therefore, we could hypothesize that the two parts draw from a total amount of energy available for the systemic process of linguistic production, balancing each other. In other words, according to our hypothesis, the phonopoiesis has a limited overall quantity of energy:

---

9 Normally, the term “phonopoietic act” indicates the complex portion of the overall communicative act that puts together verbal and non verbal aspects, the possibility to communicate by use of the vocal channel, the power to denote, evoke and connote. (Anolli & Ciceri, 1992: 97-98).
if the muscular effort (and hence the energetic expenditure) increases in the phase of vocal production and in the first part of the articulatory tract (up to the level of the back of the tongue), then force would be subtracted from the actual articulation, rendering the phonological targets less precise and the phonetic outputs weaker. The following scheme represents the just described idea of energetic balancing graphically:

![Figure 8](image)

**Figure 8.** The energetic scales.

4. **VOICE QUALITY: PARAMETERS AND DATA**

From the spectrographic morphology of our speakers’ voices emerges a systematic presence of noise, highlighted also by the parameters of aperiodicity (Jitter, Shimmer, HNR). If we compare the average of the data for each speaker to the data of a modal voice, we clearly note this constant tendency (cf. Tabb. 1, 2, 3). Here are the data found in the literature for *modal phonation* for the just mentioned parameters:

<table>
<thead>
<tr>
<th>Jitt (%)</th>
<th>Shim (%)</th>
<th>HNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.040</td>
<td>3.810</td>
<td>15.38-19.1</td>
</tr>
</tbody>
</table>

*Table 1. Modal values of the parameters of aperiodicity used.*
As already mentioned, in order to make the analysis more accurate, the same parameters were measured for the nasal and lateral consonants, for the adjacent vowels (preferential target of nasalization/denasalization processes) as well as for the entire F0 curve. As regards this last parameter, we considered the alteration of the measurements, since the Praat algorithm implements also the normal friction of the non sonorant consonantal phones in the calculus of aperiodicity. We therefore used the data measured for the entire locutionary string as a reference in order to characterize the voice quality of the speakers analyzed.

<table>
<thead>
<tr>
<th>Subject</th>
<th>F0 (Hz)</th>
<th>Jitt (%)</th>
<th>Shim (%)</th>
<th>HNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL-F</td>
<td>212</td>
<td>1.57</td>
<td>8.72</td>
<td>12.34</td>
</tr>
<tr>
<td>DM-F</td>
<td>192</td>
<td>2.5</td>
<td>11.94</td>
<td>9.82</td>
</tr>
<tr>
<td>DS-F</td>
<td>231</td>
<td>2.46</td>
<td>8.53</td>
<td>12.2</td>
</tr>
<tr>
<td>average</td>
<td>211.6</td>
<td>2.17</td>
<td>9.73</td>
<td>11.45</td>
</tr>
<tr>
<td>σ</td>
<td>19.5</td>
<td>0.52</td>
<td>1.91</td>
<td>1.41</td>
</tr>
<tr>
<td>GW-M</td>
<td>91</td>
<td>3.97</td>
<td>15.33</td>
<td>6.69</td>
</tr>
<tr>
<td>PH-M</td>
<td>114</td>
<td>2</td>
<td>12.32</td>
<td>11.38</td>
</tr>
<tr>
<td>NS-M</td>
<td>120</td>
<td>3.17</td>
<td>15.19</td>
<td>7.77</td>
</tr>
<tr>
<td>average</td>
<td>108.33</td>
<td>3.04</td>
<td>14.28</td>
<td>8.61</td>
</tr>
<tr>
<td>σ</td>
<td>15.3</td>
<td>0.99</td>
<td>1.69</td>
<td>2.45</td>
</tr>
</tbody>
</table>

*Table 2. Values measured for breath groups.*

The comparative analysis of the two tables shows a gap (actually not very strong) between the values for Jitter, Shimmer and HNR measured for the entire string and for the single sonorant segments: the reasons for this are to be found in the above made observations.
The presence, although sporadic, of *creaky voice* contributes to the deviation of the average of these parameters (compared to the standard average).

<table>
<thead>
<tr>
<th>Subject</th>
<th>F0 (Hz)</th>
<th>Jitt (%)</th>
<th>Shim (%)</th>
<th>HNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL-F</td>
<td>213</td>
<td>0,87</td>
<td>7,13</td>
<td>17,01</td>
</tr>
<tr>
<td>LM-F</td>
<td>195</td>
<td>1,74</td>
<td>11,58</td>
<td>11,84</td>
</tr>
<tr>
<td>DS-F</td>
<td>241</td>
<td>1,04</td>
<td>6,01</td>
<td>20,94</td>
</tr>
<tr>
<td>average</td>
<td>216,33</td>
<td>1,21</td>
<td>8,24</td>
<td>16,59</td>
</tr>
<tr>
<td>σ</td>
<td>23,18</td>
<td>0,46</td>
<td>2,94</td>
<td>4,56</td>
</tr>
<tr>
<td>GW-M</td>
<td>92</td>
<td>2</td>
<td>13,98</td>
<td>9,38</td>
</tr>
<tr>
<td>PH-M</td>
<td>112</td>
<td>1,3</td>
<td>11</td>
<td>13,04</td>
</tr>
<tr>
<td>NS-M</td>
<td>120</td>
<td>1,54</td>
<td>12,09</td>
<td>12,03</td>
</tr>
<tr>
<td>average</td>
<td>108</td>
<td>1,61</td>
<td>12,35</td>
<td>11,48</td>
</tr>
<tr>
<td>σ</td>
<td>14,42</td>
<td>0,35</td>
<td>1,5</td>
<td>1,9</td>
</tr>
</tbody>
</table>

*Table 3*. Values measured for hypo-nasalized segments.

The comparison between the values of the indexes of aperiodicity for hypo-nasalized and *creaky* segments shows a different mutual distribution: in particular, the rather high values for *Shimmer* and *HNR* (however close to the normal range) seem to be correlated with the velo-pharyngeal friction, differing from the values for *creaky/harsh* voice, all considerably different with reference to those of modal phonation.

Keeping in mind the above made observations, we notice a general relation between the parameters taken into account for the segments defined as hypo-nasalized. The F0 maintains modal values, similar to the average for the breath group; the *Jitter* is slightly higher than
normal (thus behaving differently compared to other amodal qualities, especially creaky voice); *Shimmer* and *HNR*, on the other hand, are the significant parameters for the effects of hypo-nasalization: high *Shimmer* (both for male and female subjects) and *HNR* values diverging from the modal threshold of about 20 dB, with sensible decrease for the males (indicating a greater distribution of energy in the noisy components of the signal).

At this point, our analysis carried out taking into account spectral and quantitative data of the indexes of aperiodicity leads us to formulate a fundamental hypothesis: the aperiodic components of the signal – to which corresponds the peculiar spectrographic structure we highlighted – can be attributed to a strong velo-pharyngeal friction and to the vibration of the velum, which, being the valve connecting the oral and the nasal tract and thus behaving as vibratile means, would imprint the relative irregularity of its movement to the aperiodic structure produced by its friction.
5. CREAKY VOICE

We already mentioned the occurrence of *creaky* vowels in the speech of the subjects analyzed (however not really significant from a
statistical point of view). It must be underlined that against our expectations, the creaky segments are more frequent in the female than in the male subjects: this fact can be audibly perceived and is confirmed by both the spectrographic analysis and the parameters measured. In fact, for the males the pitch range is constantly low (below or around 100 Hz): the male voices are basically harsh with rare creaky instances. Keep in mind that these two voice qualities essentially differ from each other as regards their frequency range, considerably below 100 Hz for creaky voice, nearly modal for harsh voice.

A certain fluctuation of F0 for creaky voice, whose tonal range found in the literature is rather strict, could be motivated by a prosodic-pragmatic movement of the tone. In fact, sometimes the spectrographic structure seems to indicate that the creaky type can occur also with F0 values slightly above those found in the literature, being however noticeably low compared to a modal pitch range.

The low fundamental frequency of this creak type of phonation is one factor that distinguishes it from harsh voice, which is otherwise somewhat similar.\(^\text{10}\)

So the most reasonable and probably most adequate label for the male speakers seems to be harsh voice, which spectro-acoustically is not very different from the creaky quality: the data of microperturbation in the period measured for our speakers coincide to a great extent with those found in the literature. Although the dividing

\(^{10}\) Laver (1980: 122).
line between *creaky* and *harsh* voice is normally said to be around 100 Hz, these two phonation types seem to be collocated along a continuum with rather homogeneous spectrographic characteristics and degree of aperiodicity of the signal. The oscillation of the F0 values for the same speaker between a minimum (maybe due to pure *creaky* vowels, concentrated in some privileged positions in the utterance\(^{11}\)) and higher values may be justified observing that, with phonatory organs in normal conditions, it is basically impossible to keep a pitch range between 24 and 52 Hz\(^{12}\). We therefore hypothesize that this is actually one single phonatory setting, despite the variation as regards F0 and the degree of microperturbation.

Another important aspect has to be interpreted in this sense: the strong aperiodic component of the signal makes the above mentioned spectrographic morphology with the little bricks less visible for the males, since it is absorbed by the more evident aperiodic velopharyngeal structure.

\(^{11}\) The positions which physiologically favour *creaky voice* are utterance-final stressed vowels (cf. Vayra, 1994), but also short pauses (stressed and unstressed) as well as the beginning of an utterance.

\(^{12}\) While the mean fundamental frequency for creak has been found to be 34.6 Hz, in an average range for male speakers of 24-52 Hz, the mean fundamental frequency for harsh voice is said to be 122.1 Hz. Laver (1980: 122).
Figure 11. Waveform and spectrogram of the phrase *It was her fairy godmother*; subject LM (F).

<table>
<thead>
<tr>
<th>Subject</th>
<th>F0 (Hz)</th>
<th>Jitter (%)</th>
<th>Shim (%)</th>
<th>HNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LL-F</td>
<td>113</td>
<td>3.53</td>
<td>15.28</td>
<td>5.52</td>
</tr>
<tr>
<td>LM-F</td>
<td>81</td>
<td>3.41</td>
<td>14.85</td>
<td>2.26</td>
</tr>
<tr>
<td>DS-F</td>
<td>107</td>
<td>2.28</td>
<td>13.02</td>
<td>5.36</td>
</tr>
<tr>
<td>average</td>
<td>100.33</td>
<td>3.07</td>
<td>14.38</td>
<td>4.38</td>
</tr>
<tr>
<td>σ</td>
<td>17.01</td>
<td>0.68</td>
<td>1.2</td>
<td>1.83</td>
</tr>
<tr>
<td>GW-M</td>
<td>89</td>
<td>6.84</td>
<td>17.87</td>
<td>3.43</td>
</tr>
<tr>
<td>PH-M</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NS-M</td>
<td>82</td>
<td>2.88</td>
<td>13.24</td>
<td>7.15</td>
</tr>
<tr>
<td>average</td>
<td>85.5</td>
<td>4.86</td>
<td>15.55</td>
<td>5.29</td>
</tr>
<tr>
<td>σ</td>
<td>4.94</td>
<td>2.8</td>
<td>3.27</td>
<td>2.63</td>
</tr>
</tbody>
</table>

Table 4. Values for *creaky* segments.

It seems possible to extend the general definition of *lax voice*, which emerged from the description of the overall setting of the phonation-articulation typical in *Scouse*, to cover a series of factors of vocal emission which partly go beyond the simple laryngeal
mechanism (rhythm, intensity, aspects of hypo-articulation) and – what is fundamental for our analysis – the supra-laryngeal phonatory dynamics. As is highlighted by Ní Chasaide & Gobl (1997: 451): *for all voice qualities, the reader should bear in mind that they are not fixed entities. Non-modal qualities may occur to a greater or lesser degree, i.e., may be further from or closer to modal voice. Voice qualities can also be of a compound type, as for example in whispery creaky voice.*

Moreover, every voice quality can vary considerably during the utterance and the *creaky* voice is particularly sensitive to alterations of this kind. So, if we consider the superimposition of acoustic effects caused by the synergic mechanisms which operate in contiguous parts of the phonatory system, using a finite quantity of energy not equally distributed, we can outline a quite clear perspective of the overall characterization of the phonopoietic act in *Scouse*.

In the table below, we present the number of *creaky* segments in proportion to the number of breath groups analyzed. As already mentioned, the data show a higher incidence in the female subjects, while speaker PH produced no pure *creaky* segments at all, just due to the above mentioned peculiar vocal connotation of the male subjects.

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nr. creaky segm. per B.G.</strong></td>
<td>10/11</td>
<td>12/17</td>
</tr>
<tr>
<td></td>
<td>4/11</td>
<td>5/11</td>
</tr>
<tr>
<td></td>
<td>5/14</td>
<td>0/13</td>
</tr>
<tr>
<td><strong>Average per B.G.</strong></td>
<td>1</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>0.38</td>
<td>0</td>
</tr>
</tbody>
</table>

*Table 5.* Distribution *creaky* segments - hypo-nasalized segments per breath group.
According to the different modalities of muscular and postural activation of the laryngeal mechanism (on the vertical and horizontal level, like schematized above), we can define the different non-modal phonatory settings, which, with regard to the laryngeal features\textsuperscript{13}, correspond to:

a) \textit{Creaky voice} \sim [++ \text{slack vocal cords}; --\text{spread glottis}]

b) \textit{Lax voice} (\approx - \text{tense voice})\textsuperscript{14} \sim [++ \text{slack vocal cords}; --\text{spread glottis}]

With reference to this scheme, also on the laryngeal level there is no impediment to the interpretation of a phonatory compound which reconciles \textit{creaky} and \textit{lax} voice, as seems to be provable for \textit{Scouse}.

\textsuperscript{13} This annotation (++/+/-/-) indicates the degree of activation of the single segment. Cf. Halle & Stevens (1971).

\textsuperscript{14} Ní Chasaide & Gobl (1997: 443-452).
6. VOICE QUALITY AND GENDER

In our corpus, there seems to be a general correlation between voice quality and gender. In fact, the detailed description of the Scouse phonatory setting allows us to identify a certain differentiation in the vocal behaviour of the male and female subjects analyzed. According to the parameters and the acoustic analysis we can notice the following facts:

a) for the females we find a voice quality perceived as “nasal” with normal pitch values and rather frequent instances of creaky voice in correspondence with the stressed segments (mainly in final position or near pauses in the breath group);

b) for the males the voice quality perceived is harsh. The characteristics of this phonatory setting dominates over the impression of “nasality”, which is much less perceptible compared with the female subjects; the average pitch values are constantly low (about 100 Hz), while creaky segments are rare and even absent for one speaker (PH). However, keep in mind the above made observations (cf. § 5) for a possible activation of a continuous harsh/creaky phonatory compound for the male subjects.

All things considered, it seems to be possible to observe a gender divergence for the overall definition of the Scouse voice quality. Limited to the perceptive level, we can identify a different connotation for the female voice (which uses a high range of frequency and a high volume, denoting the “nasal” effect as salient colour feature)
compared to the male one (characterized by a strongly compressed tonal range, an extremely low volume and a colour that indicates perceptive elements definable as “nasal” on an essentially harsh basis.

7. CONCLUSION

From a critical review of the literature on the Scouse voice quality emerge some inaccuracies. Former descriptions explain in detail the articulatory characteristics of the variety of English spoken in Liverpool, but they seem not to examine in depth the phonatory dynamics which are the basis of an acoustic perception commonly recognized as peculiar and marked. On the basis of the spectro-acoustic analysis carried out we propose the definition “hypo-nasalized” for the Scouse vocal setting due to the convergence of measurements, analysis of the spectrographic morphology and of the perceptible vocal colour. The dominant factor seems to be the strong velo-pharyngeal friction; hence the vocal output is distorted since it is influenced by a constant feature of limited nasal resonance. This feature is understood as characterizing the voice quality, being diffused among the non nasal phones, while the nasal ones undergo a lowering of the nasal resonance due to the partial lowering of the velum. There seems to be a general correlation between gender and voice quality in terms of the above mentioned marked characterization.

The basic hypothesis is that the supra-laryngeal phonatory setting, which is expensive in energetic terms, is balanced by the articulatory
setting definable as *lax*. So, the systemic perspective we presented can be defined as “energetic scales”.

**Massimiliano Barbera & Marlen Barth**  
University of Pisa  
m.barbera@ling.unipi.it, marlen_barth@yahoo.com

*Although the paper has been conceived and developed by both Authors, for academic reasons, the responsibilities should be divided as follows: M. Barbera: § 3-4-5-6; M. Barth: § 1-2. The conclusions (§ 7) are by both Authors.*
REFERENCES


