Variation of glottal LF parameters across F0, vowels and phonetic environment

Michelle Tooher & John McKenna
School of Computing, Dublin City University
Context

• Machine learn characteristics of a speaker
  – Given utterance info. (prosodic, contextual and info about individual utterance)
  – Predict LF parameters of glottal flow for speaker and utterance
Data

- 2 male speakers
- 3 vowels - /a/ /i/ /u/
- 4 contexts - /s_t/ /s_d/ /z_t/ /z_d/
- “Say ___ again”
- 7 pitches: 90 – 210
- Randomly presented
- 3 sets
Analysis & Fitting

• Kalman-Filter based LP (McKenna, ‘99)
  – Chooses closed phase sections
  – Performs closed phase covariance LP
  – DGF

• LF fitting (Fant et. al., ‘85)
  – LF parameters: \( t_e, t_c, t_p, T_a \)
LF model
Questions

• Does glottal flow vary w.r.t. utterance and speaker?
• Any distinct patterns?
• What influences these variations/patterns?
• Should they be taken into consideration?
• Speaker specific?
Data Analysis

• LF parameters from beginning, middle, end of each vowel
• Statistical analysis (SPSS)
• Data plots
Data Analysis

- SPSS – correlation analysis*

<table>
<thead>
<tr>
<th></th>
<th>$T_0$</th>
<th>$t_p$</th>
<th>$t_e$</th>
<th>$t_c$</th>
<th>$T_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_0$</td>
<td></td>
<td>.757</td>
<td>.902</td>
<td>.964</td>
<td>.446</td>
</tr>
<tr>
<td>$t_p$</td>
<td></td>
<td></td>
<td>.949</td>
<td>.859</td>
<td>-.065</td>
</tr>
<tr>
<td>$t_e$</td>
<td></td>
<td></td>
<td></td>
<td>.961</td>
<td>.132</td>
</tr>
<tr>
<td>$t_c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.364</td>
</tr>
<tr>
<td>$T_a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$T_0$</th>
<th>$t_p$</th>
<th>$t_e$</th>
<th>$t_c$</th>
<th>$T_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_p$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_e$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T_a$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Pearson Correlation Coefficients
Results

• Variations w.r.t. $T_0$:
  
  - $t_c$, $t_e$, and $t_p$ rise
  
  • $t_c$ and $t_e$ close to linear whereas $t_p$ portrays nonlinearity

  - $T_a$ - little variation
Results

/u/

\begin{center}
\begin{tikzpicture}
\begin{axis}[
    xlabel={$T_0$},
    ylabel={$t_p$},
    xmin=4, xmax=12, xtick={4,5,6,7,8,9,10,11},
    ymin=0, ymax=6e3, ytick={0,1,2,3,4,5,6},
    xtick scale label code/.code={},
    ytick scale label code/.code={},
    xticklabel style={font=\footnotesize},
    yticklabel style={font=\footnotesize},
    width=\textwidth,
]
\addplot[red, mark=x, only marks] table [col sep=comma] {data1.csv};
\addplot[blue, mark=x, only marks] table [col sep=comma] {data2.csv};
\addplot[red, dotted] table [col sep=comma] {line1.csv};
\addplot[blue, dotted] table [col sep=comma] {line2.csv};
\end{axis}
\end{tikzpicture}
\end{center}

27/08/03
VOQUAL '03
Results (cont.)

• Variations w.r.t. vowels:
  - LF parameter values of /a/ higher than /i/, and /u/
  - Linear regression shows significant differences in both slope and y-intercept between /a/ and /i/ or /u/
• Variations w.r.t. environment:
  – Both linear regression and data plots show the following:
    • /z/ preceding – affects parameters
    • /s/ preceding – no apparent effects
    • Context following vowel has no effect on parameters
    • Voiced and voiceless pairs /s_t/, /z_d/ - no effect
Results (cont.)

• Variations in waveshape parameters
  – $R_a$ and $R_k$ change with $T_0$
  – $R_g$ varies little
  – As $T_0$ rises $R_d$ also rises

• Variations w.r.t speaker
  – Same patterns across two speakers
  – S2 values are generally higher than S1
Conclusions

• Patterns do exist
• LF parameters vary with $T_0$
• $t_c$ and $t_e$ appear to vary linearly with $T_0$ whereas $t_p$ and $T_a$ appear to vary non-linearly
• Only the voiced context appears to have an affect on the parameters and only when it preceeds the parameters
Conclusions

• Vowel influences values of parameters

• Variations with $T_0$ are not speaker specific, however values of the LF parameters are

• More data across speakers, contexts and vowels is needed for a more exhaustive study
Question

• How could this affect synthesis?
  – F0 manipulation – parameters need to be adjusted but possibly at different rates
  – Original and target environments
  – New speakers – can a new speaker be created by varying the levels (y-intercept) of the parameters?
• Additional data plots follow