SSB-CARRIER MISMATCH DETECTION FROM SPEECH CHARACTERISTICS: EXTENSION BEYOND THE RANGE OF UNIQUENESS

Thomas Gülzow, Ulrich Heute
Inst. f. Circuit & System Theory, Fac. of Engineering,
Univ. of Kiel, Kaiserstr. 2,
D-24143 Kiel, Germany
Tel: +49 431 880 6125; fax: +49 431 880 6128
e-mail: uh@tf.uni-kiel.de

Hans J. Kolb
MEDAV GmbH,
Gräfenberger Str. 34,
D-91080 Uttenreuth, Germany

ABSTRACT

In certain radio-surveillance applications, speech signals with uncertain carrier frequencies may be detected. Beyond other disturbances, carrier mismatches distort the demodulated SSB signals by frequency shifts Δf.
Recently, a detection and correction of such errors was presented [1], exploiting the harmonic structure of voiced speech with a fundamental (or "pitch") frequency f_p. Noise-robust, modified pitch-detection methods [2, 3] were shown to be applicable, together with an iteratively corrected "coherent- superposition" algorithm [4, 5]. Thereby, carrier errors are removed reliably [1] though only within an a-priori *uniqueness range* $|\Delta f| < f_p/2$.
In the following, the extension beyond this limitation is presented. It emerges from a closer, statistical analysis of the measurement errors: The measurement statistics within this range are interpreted as a conditional probability density function (pdf) $p_{\Delta f}(x|y)$ conditioned on the actual pitch frequency, its combination with the (measurable) pitch-frequency pdf $p_{f_p}(y)$ yields the joint pdf $p_{\Delta f, f_p}(x, y)$, and its integration gives the marginal pdf $p_{\Delta f}(x)$. Its evaluation leads to an even narrower reliability range $|\Delta f| < f_p/3$, first. The interpretation of the uniqueness range as one period of an f_p-periodic repetition of the measurement statistics, however, leads to the solution: After the same steps as sketched above, the marginal pdf $p_{\Delta f}(x)$ shows a strong peak at the true value Δf without a (theoretical) restriction. The resulting algorithm was successfully tested with simulated and real data.

1 INTRODUCTION

In certain short-wave radio-channel surveillance scenarios, speech signals from non-co-operative sending stations are detected at some uncertain carrier frequencies. Beyond fading, additive noise, and strongly amplified disturbances in speech pauses, carrier mismatches leave distortions due to a frequency shift Δf of the demodulated SSB signals. This is audible for $|\Delta f| > 5Hz$ for noise-free signals and careful listening, it is more distinct at 10 Hz, and it becomes annoying even when covered by additive noise at 15 Hz and more.

Recently [1], an approach to the detection and correction of carrier errors was presented. It exploits the fact that the harmonic structure of voiced speech sections is destroyed by a shift, whereas a regular line spectrum, equally spaced by the so-called pitch frequency f_p, is maintained. So, the well-known, large collection of pitch-detection methods [6] should offer a solution; however, a method is needed which is both robust to the strong additive noise [2, 3] and modified to be able and handle the frequency shift [4]. Two versions were presented in [1], one based on an idea similar to "vector-quantization", one using Atal's "coherent superposition" of cosines [5] with an "iterative correction" [4], and especially the latter one was shown to yield reliable detections of carrier errors as long as $|\Delta f|$ is not too large.

It can be seen easily from the line-frequency pattern of a periodic signal with fundamental frequency f_p, that a shift by $+f_p/2$ and by $-f_p/2$ yield identical line patterns (except perhaps for the line at zero frequency, which, however, may be missing in the measurement, as also any other line cannot be guaranteed to be found!). Vice-versa, it is impossible to find the correct sign of Δf from a corresponding pattern. Due to measurement uncertainties, already patterns close to the above situation become more and more unreliable (see [1]), and cases with $|\Delta f| > f_p/2$ are mapped into this "uniquenesss interval", if the detection relies on the iterative correction towards the closest harmonic grid.

In the following, a statistical analysis of the Δf measurement is carried out. It explains the above observation by a pdf limited to the range $|\Delta f| < f_p/2$. While the qualitative error behaviour can be seen here already, a formal description is achievable via some additional considerations: Obviously, it depends on the

actual value of the fundamental frequency f_p. Writing the above pdf as a conditional pdf and multiplying it by the pdf of the pitch frequency itself gives the bi-variate pdf $p_{\Delta f, f_p}(x, y)$ of general $(\Delta f, f_p)$ pairs, from which an integration over $y \sim f_p$ yields the (marginal) pdf of the measurement Δf.

This derivation is exposed in section 2, together with an evaluation in terms of gross errors ("outliers" with errors above 20 Hz). An even narrower usability range $|\Delta f| < f_p/3$ results. In section 3, however, a different interpretation of the range-limited measurement pdf is introduced, leading to its periodic repetition with a period f_p. Starting from here and following the same steps as above yields a marginal pdf finally which does contain a strong maximum at the correct value Δf without a range restriction. An algorithm using this result and exploiting random variations of the pitch frequency in consecutive voiced speech segments is validated in section 4. Section 5 contains a summary and conclusions.

2 Basic Algorithm and Statistical Analysis

Robust Pitch Detector: Voiced-speech signal blocks are (quasi-) periodic, therefore characterised by integer multiples of a fundamental frequency, usually termed pitch frequency f_p , which is in the order of 60 Hz (male, low) to 300 Hz (female, high, or child). Many pitch-detection algorithms [6] were developed in the past, mainly in the context of speech coding. Especially for an application in forensic pitch-contour determination from very noisy signals, an approach was worked out [2, 3] which is based on two main ideas:

- A high-order ($n > 20$) linear prediction yields a polynomial A(z), whose zeros include some parasitic points, but also at least some of the harmonics of f_p as their position angles in the z plane. By means of a simple ("Bistritz-") test, it is possible to find the "relevant" zeros with low computational effort, since they are close to the unit circle $z = e^{j\Omega}$.

- The first zero with $\Omega > 0$ should give f_p directly, as its angle; however, the zeros are not immediately evaluated as such. Instead, following an idea of Atal [5], cosines of all detected frequencies are simply added. For an infinite number of correct harmonics, this "coherent superposition" should yield an impulse train with pulses separated by $1/f_p$. For at least some (...3...) roughly correct harmonics, still "smoothed" pulses with low-error distances result (see Fig. 1), and with some additional refinements, this was shown to work even at an SNR = 0 dB [2, 3].

Problem: If the speech signal contains a frequency shift, however, the superposition does not add cosines of frequencies $\lambda \cdot (f_p + \Delta f), \lambda \in \mathbf{Z}$, but components at $(\Delta f + \lambda \cdot f_p)$, which does no more yield a pulse-like sum with a simple relation to the shifted fundamental frequency.

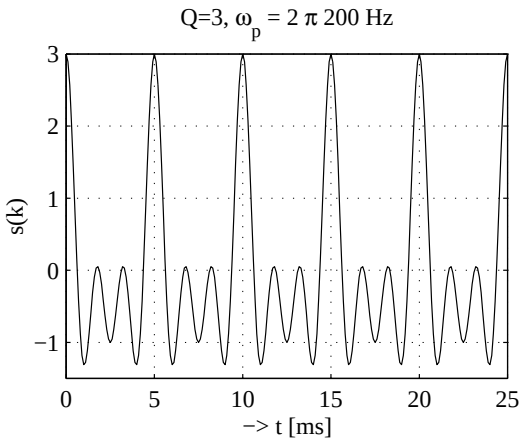

Figure 1: Coherent superposition of 3 harmonics.

Solution: Of the two approaches discussed in [1], the following one is preferable: If, assuming $\Delta f > 0$, after a first superposition a second one with slightly *smaller* frequencies is calculated, the sum signal becomes "more pulse-like"–i.e., the second maximum value increases and approaches the first maximum. Else, the second maximum is reduced–so the test has to be repeated with a set of slightly *higher* frequencies. An iteration results in the correct value f_p , and the sum of the changes gives the shift Δf.

Measurement Statistics: Fig. 2 depicts a histogram of Δf estimations measured from simulations with $\Delta f = 0$. According to [4], this may be modelled as a superposition of 3 GAUSSian components. (For the sake of simplicity, only one term will be written and sketched explicitly, in the following.) As can be seen easily from

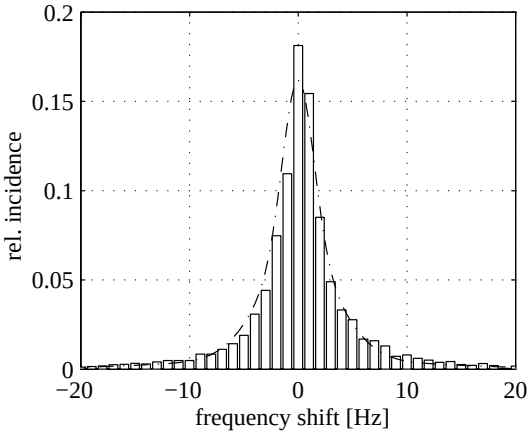

Figure 2: Histogram of Δf estimations for a true $\Delta f = 0$ and approximating sum of 3 GAUSS pdf's.

the line-frequency pattern of a periodic signal with fun-

damental frequency f_p, shifts by $+f_p/2$ and by $-f_p/2$ yield identical zero patterns in the above analysis. So, for a true shift $\Delta f \neq 0$, the pdf of Fig. 2 is not only shifted by Δf , but also "aliased" with its tails into the range $|\Delta f| < f_p/2$ (see Fig. 3). It is obvious from this schematic that there is a growing probability for a wrong estimation $\Delta f' = \Delta f \pm f_p$, if the limits of the uniqueness range are approached or even exceeded. As f_p itself is

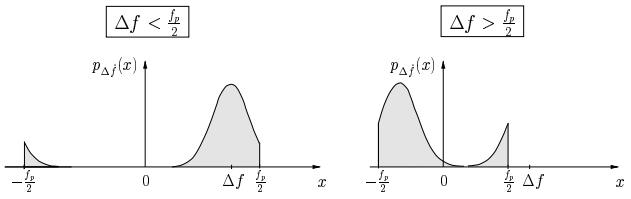

Figure 3: Estimation pdf's with aliasing into the range $|\Delta f| < f_p/2$.

varying with time (even for one single speaker) and may be described by a pdf $p_{f_p}(y)$, the above sketched pdf actually is a conditional pdf $p_{\Delta f}(x|y)$ conditioned on f_p. From measurements on a large set of female and male speakers, it was found that f_p variations around the individual average pitch frequencies can be described by a LAPLACE pdf $p_{f_p}(y)$ (see Fig. 4). Then, the joint pdf of the shift and pitch frequency variables is found, finally, as the product (where the internal summation is due to the "mirrored" tails of Fig. 3)

$$\begin{aligned} p_{\Delta f,f_p}(x,y) &= p_{\Delta f}(x|y) \cdot p_{f_p}(y) = \\ &\frac{1}{\sqrt{2\pi}\sigma_{\Delta f}} \cdot \sum_{k=-\infty}^{\infty} exp(-\frac{(x-\Delta f - k\cdot f_p)^2}{2\sigma_{\Delta f}^2}) \\ &\cdot \frac{1}{\sqrt{2}\sigma_{f_p}} \cdot exp(-\frac{|y-\mu|}{\sigma_{f_p}}), \quad |x| < f_p/2, \\ p_{\Delta f,f_p}(x,y) &= 0, \qquad\qquad |x| \geq f_p/2, \end{aligned} \quad (1)$$

as to be seen in the 3-D plot of Fig. 6.a). From Eq. (1), an integration over $y \sim f_p$ yields the marginal pdf

$$p_{\Delta f}(x) = \int_{y=-\infty}^{\infty} p_{\Delta f,f_p}(x,y)dy, \quad (2)$$

where the integration can be carried out numerically over a finite interval, because of the physical limits of possible pitch frequencies. This pdf describes the error probability for a fixed true shift Δf, as depicted in Fig. 5. It becomes visible again that there are gross errors with increasing Δf, and that no reasonable result is obtained if Δf is close to or above half the speaker's average pitch frequency. Defining "gross errors" as measurements deviating by at least 20 Hz from the true shift Δf, we find a reliability range even smaller than the uniqueness range: The approach is limited to roughly $\Delta f \in (-f_p/3, +f_p/3)$.

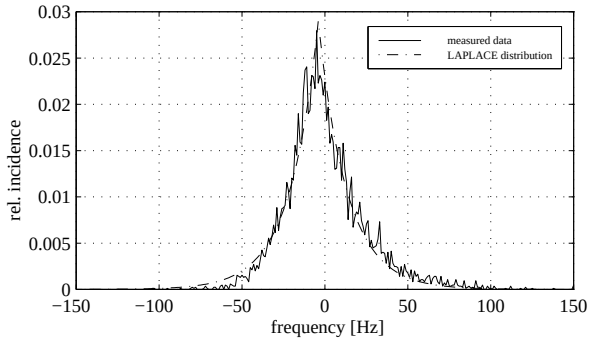

Figure 4: Histogram and LAPLACE pdf for pitch variations of various speakers after subtraction of individual average pitch frequencies.

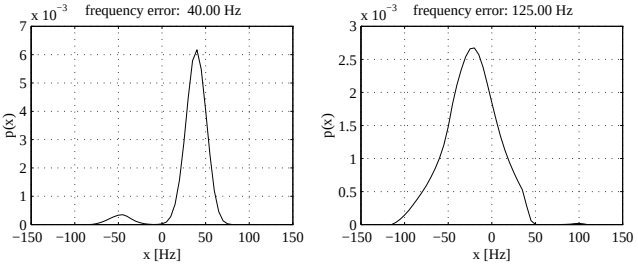

Figure 5: Measurement pdf's for $\Delta f = 40Hz$ *and* $\Delta f = 125Hz$*, assuming a speaker with an average pitch* $f_p = 150Hz$*, varying according to a LAPLACE pdf with* $\sigma_{f_p} = 24Hz$.

3 Periodic PDF Interpretation and Range Extension

The inversion of the above-sketched chain of arguments shows a way to overcome these narrow restrictions, which would be very severe especially for low-pitch male speakers: The reason for the original limitation to $|\Delta f| < f_p/2$ was the fact that a measured value Δf is not unique–it may as well be replaced by $\Delta f \pm f_p$. The same holds for any multiple of f_p. This means, however, that Fig. 2 can be interpreted as just one period of an *f_p-periodic pdf* $p_{\Delta f}(x|y)$–and f_p is known as a side-result of the algorithm. Inserting this function into Eq. (1) leads to Fig. 6.b): The "mountain" of Fig. 6.a) is replaced by "many mountains" along star-formed lines following the increasing pitch frequencies' harmonics– and now: one of these "mountains" is situated on the exact frequency shift value Δf. Therefore, the integration of Eq. (2) applied here again, leads to a summation of (near-)maxima only at Δf, while for all other points x random values add up: The resulting

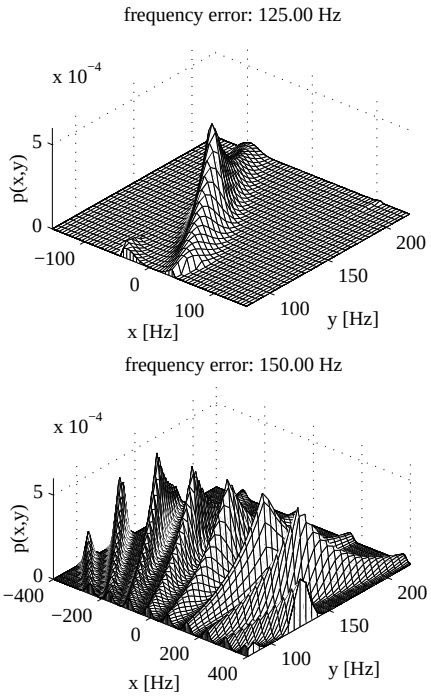

Figure 6: Bivariate pdf's $p_{\Delta f, f_p}(x, y)$ assuming the same speaker parameters as in Fig. 5 and a) $\Delta f = 125Hz$, with range limitation as in Eq. (1), b) $\Delta f = 150Hz$, with periodic interpretation for $|x| > f_p/2$.

marginal pdf shows a distinct maximum at the correct frequency shift, as can be verified from Fig. 7.

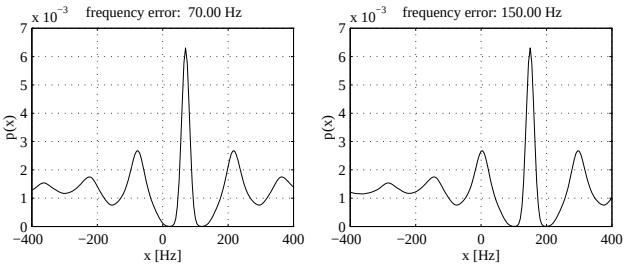

Figure 7: Estimation pdf $p_{\Delta f}(x)$ based on a periodically repeated conditional pdf (with parameters as in Figs. 5 and 6), a) $\Delta f = 70Hz$, b) $\Delta f = 150Hz$.

4 Results

An algorithm following the above lines, namely, pitch-detection via high-order prediction, harmonic superposition with iterating correction, periodic interpretation of the measurement pdf and evaluation of a pdf, i.e. in practice, a histogram over L consecutive frames of voiced speech, is described in detail in [4]. It was found from simulations both with artificial and real data that L = 30 voiced segments of length 40 msec each suffice to estimate $\Delta f \in (-300, 300)Hz$ with a gross-error rate below 8 % even for an SNR = 0 dB with both a bias and and rms error below audibility (see Fig. 8).

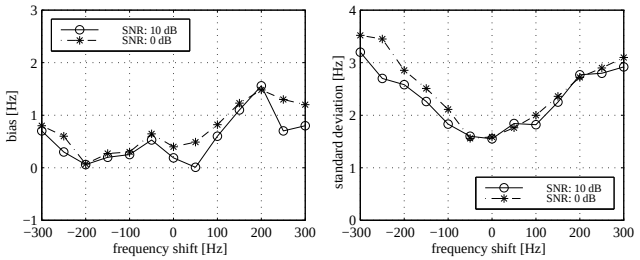

Figure 8: Bias and standard deviation of Δf-measurements in dependence of the true shift Δf.

5 Concluding Remarks

The pitch-estimation based method for the detection of a carrier-frequency mismatch of SSB-demodulated speech signals proposed in [1] was extended beyond the uniqueness range $|\Delta f| < f_p/2$ (and the even smaller reliability range range $|\Delta f| < f_p/3$ resulting from a statistical analysis) by exploiting just the non-uniqueness feature. The remaining limitation to a range $|\Delta f| < 300Hz$, visible in Fig. 8, is a practical one, stemming from some details of the realisation in [4], especially an analysis-band limitation to roughly 1200 Hz for reasons of computation time.

Above, it was observed that a histogram over $L = 30$ voiced segments suffices for a reliable detection. The corresponding signal duration of 1.2 sec means that even slowly drifting carrier frequencies can be tracked by the same method.

REFERENCES

[1] Gülzow, Th., Heute, U., Hossen, A.N., Kolb, H.J.: Detection and Correction of SSB Carrier-Frequency Mismatch by means of Speech-Signal Characterstics. Submitted to: EURASIP-IEEE Int. Conf. BIOSIGNAL, Brno, Czech Rep., 2002.

[2] Arévalo, L.: Contributions to the Estimation of Frequencies for Noisy, Short-Duration Oscillations and an Application to Speech-Signal Analysis (in German). Dissertation, AGDSV / Ruhr-Univ., Bochum, Germany, 1991.

[3] Arévalo, L.: Linear-Predictive Eigenvalue-Oriented Pitch-Contour Measurement. Proc. 5-th ASSP Workshop Spectr. Estim. Model., pp. 299-303, Rochester, N.Y., USA, 1990.

[4] Gülzow, Th.: Quality Enhancement for Severely Disturbed Speech Signals: Carrier-Shift Detection and Additive-Noise Suppression (in German). Dissertation, LNS/TF/Univ. Kiel, Germany, 2000.

[5] Atal, B: Speaker Recognition Based on Pitch Contours. JASA, vol. 52 (1972), pp. 1687-1697.

[6] Hess, W.: Pitch Detection. Springer, Berlin New York, 1983.