



SYNTÉZA ŘEČI

Petr Horák

horak@ufe.cz

Ústav fotoniky a elektroniky, v.v.i.
AV ČR, Praha

Obsah

- Úvod a historie
- Zpracování textu
- Modelování prozodie
- Metody syntézy řeči
- Aplikace syntézy řeči

Komunikace mezi lidmi

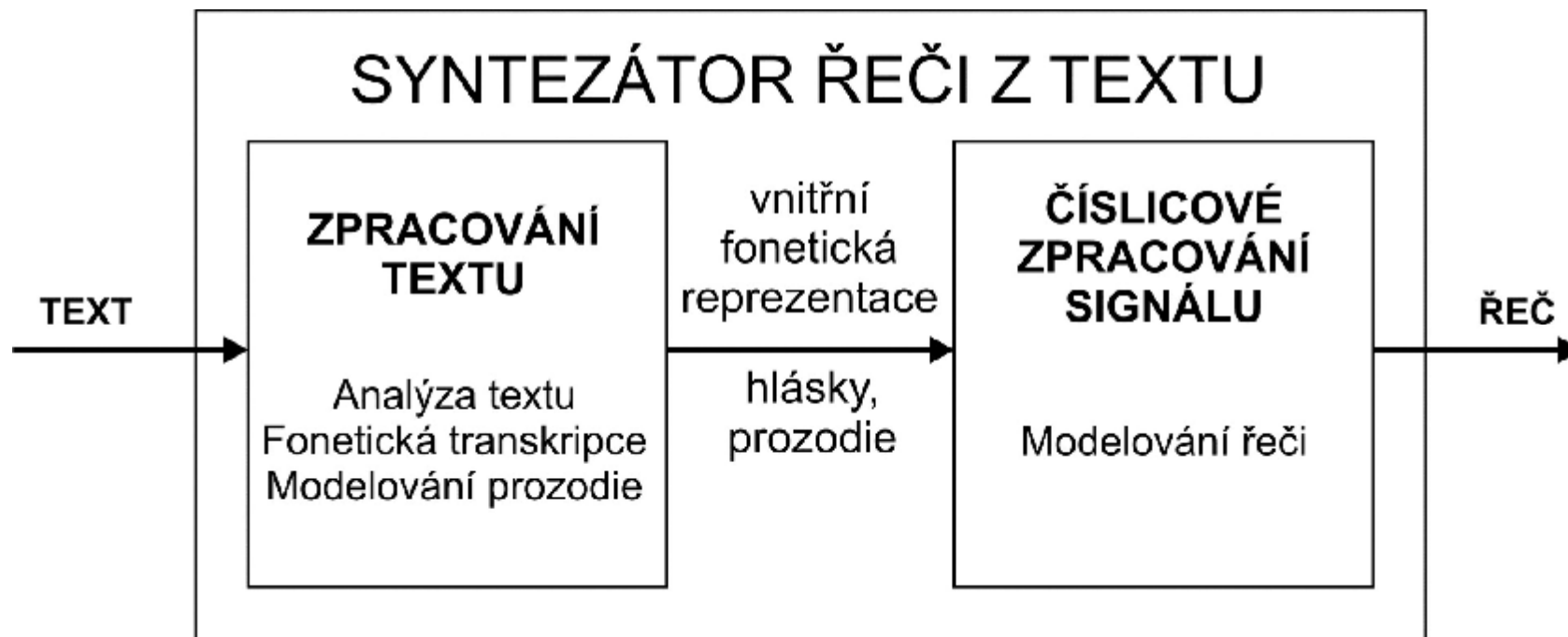
- **Písmo – psaná podoba komunikace**
 - věty, slova, písmena
- **Řeč – mluvená podoba komunikace**
 - **akustika**
 - vytváření a vnímání řeči
 - akustické vlastnosti řeči (formanty)
 - **fonetika a fonologie**
 - **lingvistika**
 - **prozodie**
- **Fonetická informace**
- **Prozodická informace**

Syntéza řeči z psaného textu

- **Řeč – nejpřirozenější forma komunikace**
- **Syntéza řeči – proces umělého vytváření řeči**
- **Cíl: vyvářet řeč v takové formě a kvalitě, aby nebyla rozpoznatelná od přirozené řeči**

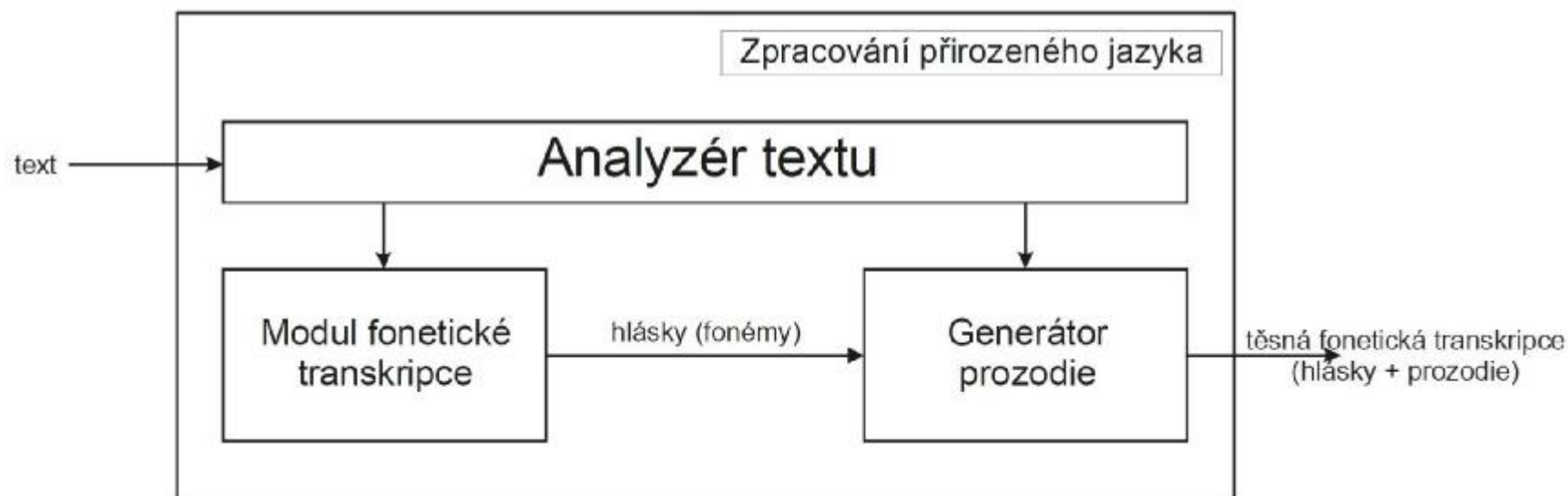
Syntéza řeči z psaného textu

- **Nejobecnější úloha: na vstupu text, na výstupu řeč**
- **Dělí se na dva základní moduly**
 - modul pro zpracování textu
 - řečový syntetizér



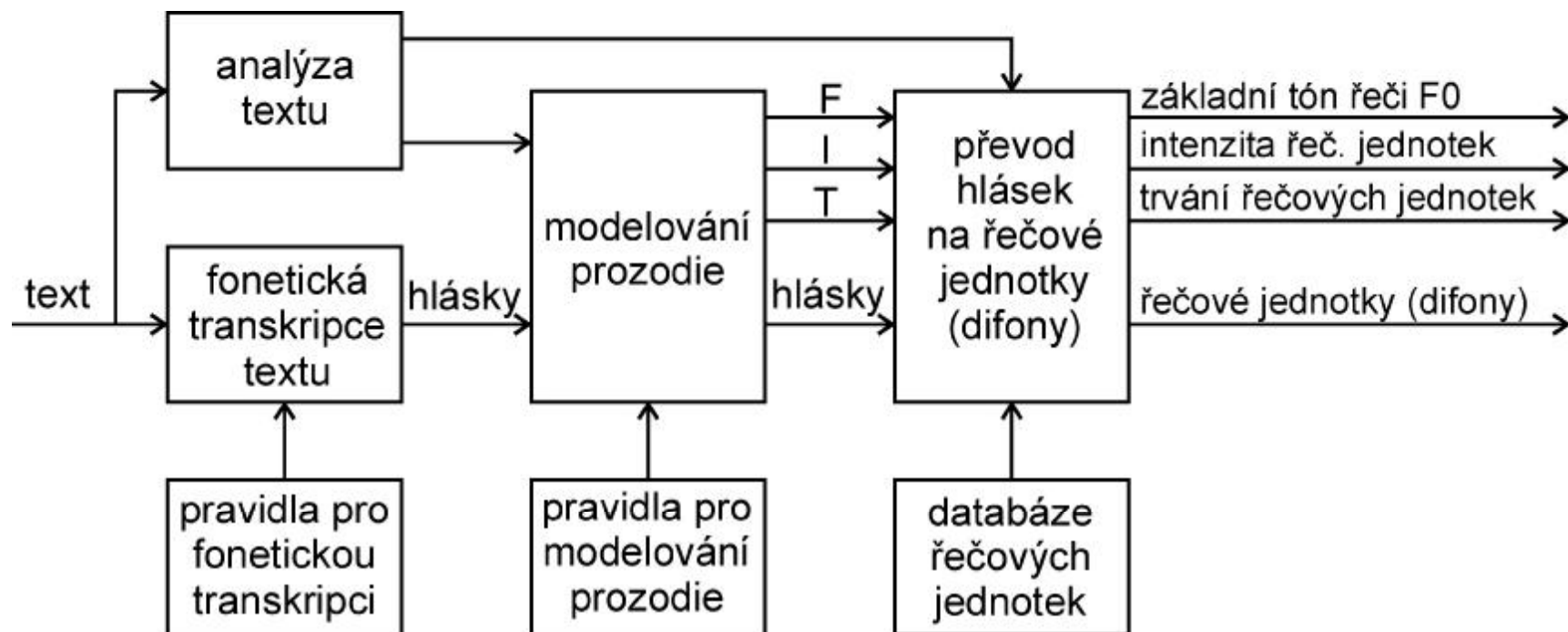
Zpracování textu

- **Zpracování textu = zpracování přirozeného jazyka**
(Natural Language Processing, NLP)
 - **Analýza textu**
 - **Fonetická transkripce**
 - **Generování prozodických charakteristik**

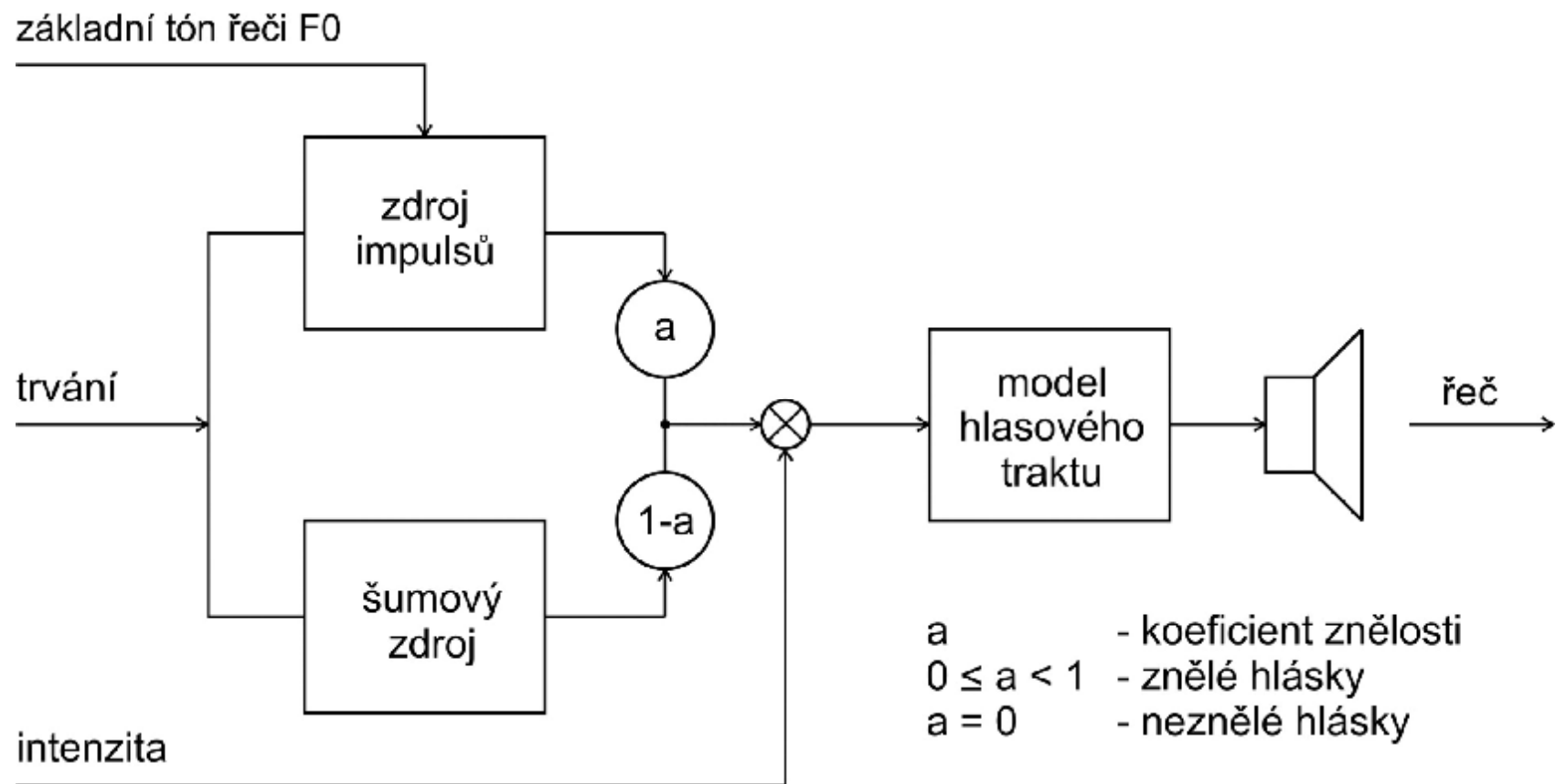


Zpracování textu

- Příklad modulu pro zpracování textu



Řečový syntetizér

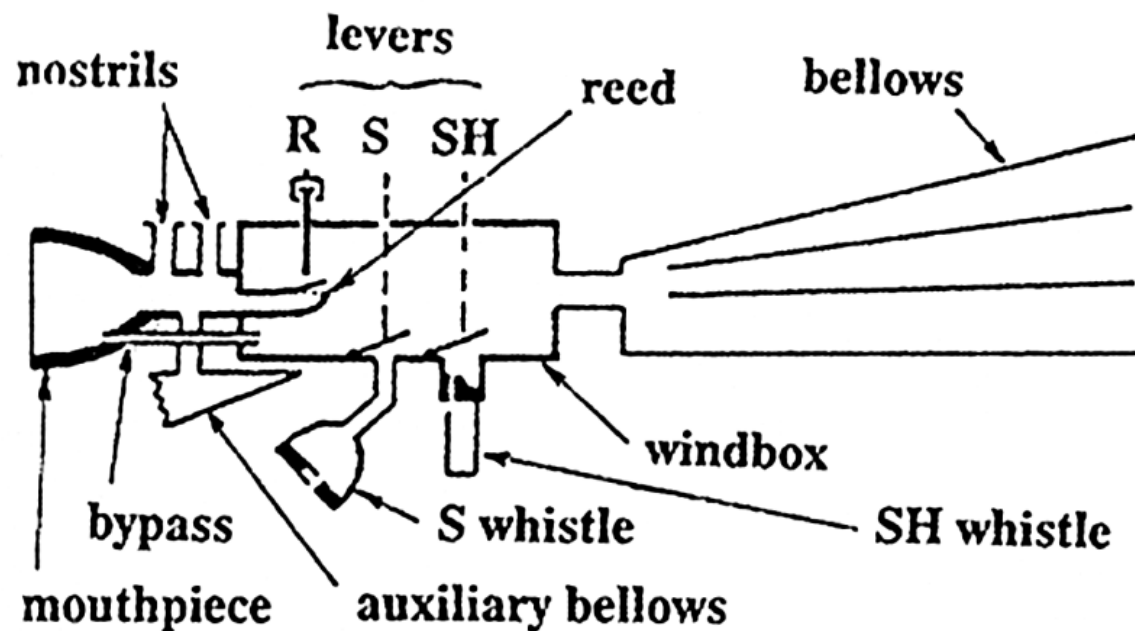


Historie syntézy řeči

- **Mechanické syntetizéry**
 - Kratzenstein (1779) (první pokusy – samohlásky)
 - Kempelen (1791) (mluvící stroj)
- **Elektronické syntetizéry**
 - Stewart (1922) (dva rezonátory)
 - Dudley (1939) (Voder – první syntetizér souvislé řeči)
 - Fant (1953) (OVE I – kaskádní formantový syntetizér)
- **Digitální syntetizéry**
 - Klatt (1979) (MITalk, Klattalk – nejlepší formantová syntéza)

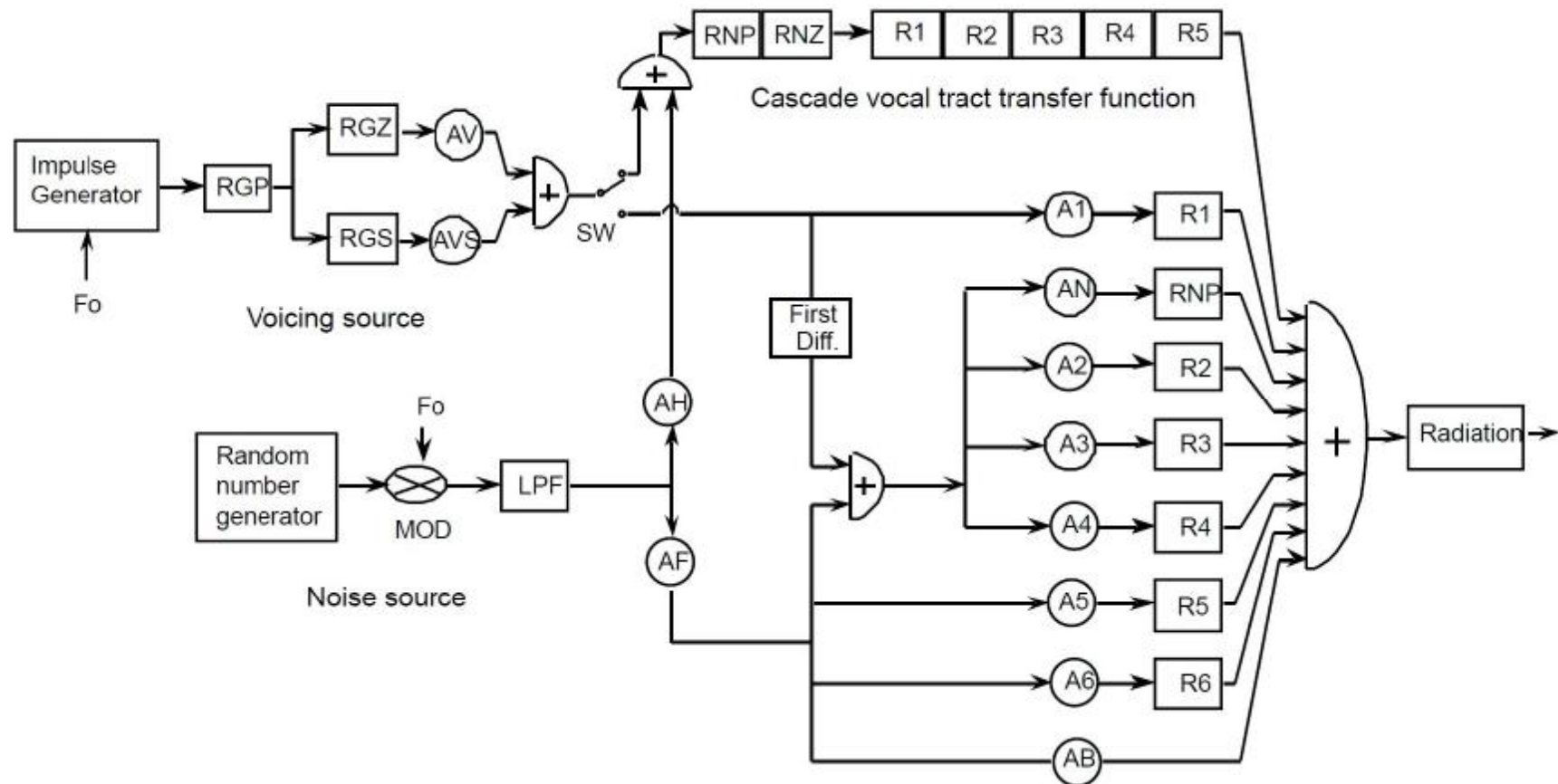
Kempelenův syntezátor

- Jednalo se o mechanické zařízení napodobující artikulační orgány člověka



Klatt

- Blokové schéma Klattova syntezátoru

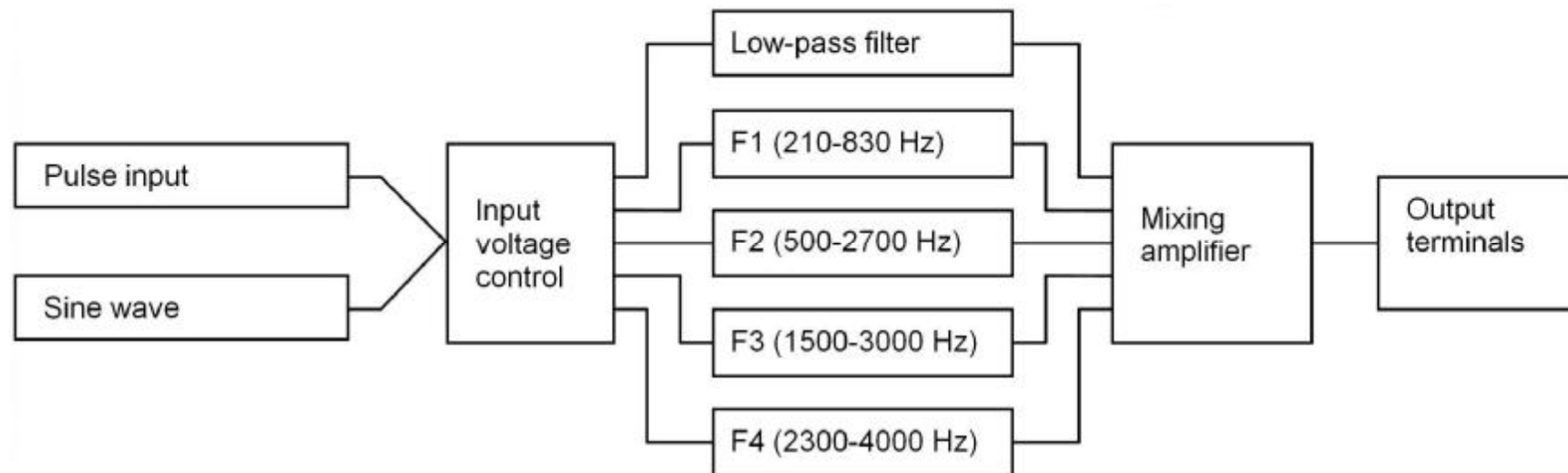


Historie české syntézy řeči

- **Kaňka**
 - první pokusy (20. léta 20. století)
- **Janota**
 - první český syntetizér řeči (1964)
- **Ptáček, Borovičková, Maláč**
 - OVED 1 – první syntetizér souvislé řeči (1968)
 - HO2 – první český TTS systém (1977)
- **ÚRE AV ČR, Ptáček**
 - PCVOX – první český LPC TTS systém (1990)
 - Epos – první český open source TTS (1996)
- **ZČU Plzeň**
 - ARTIC – první český Unit Selection TTS (2000)

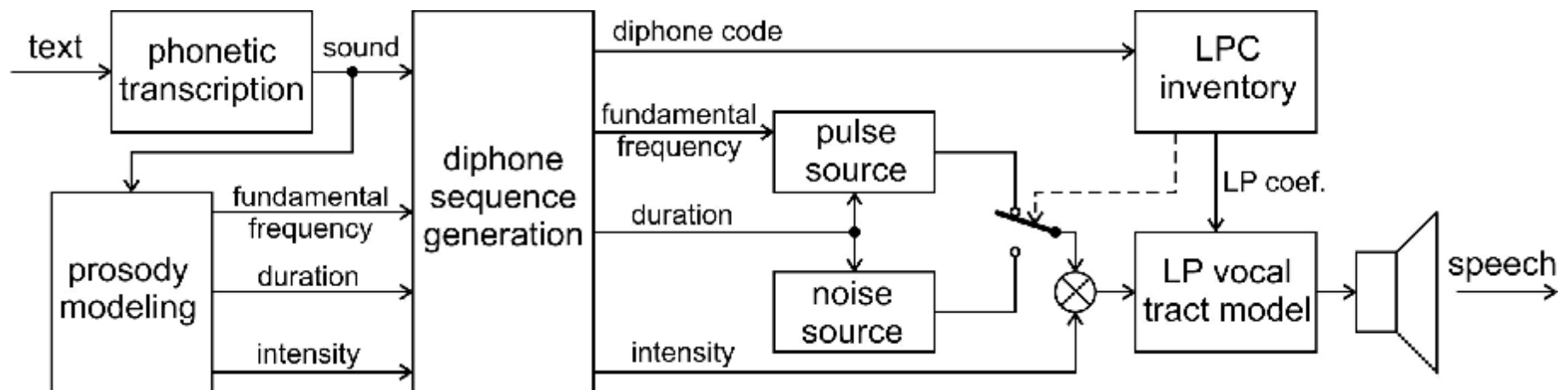
Janota (1964)

- Jednoduchý paralelní formantový syntetizér češtiny



LPC TTS - Vích, Ptáček (1990-1992)

- První český LPC TTS systém (1990-1992)



LPC TTS - Vích, Ptáček (1990-1992)

- **Komerční podoba českého LPC systému pro převod psaného textu na řeč**
 - **Vyráběno 1992 až 2000 Výrobním družstvem nevidomých SPEKTRA**



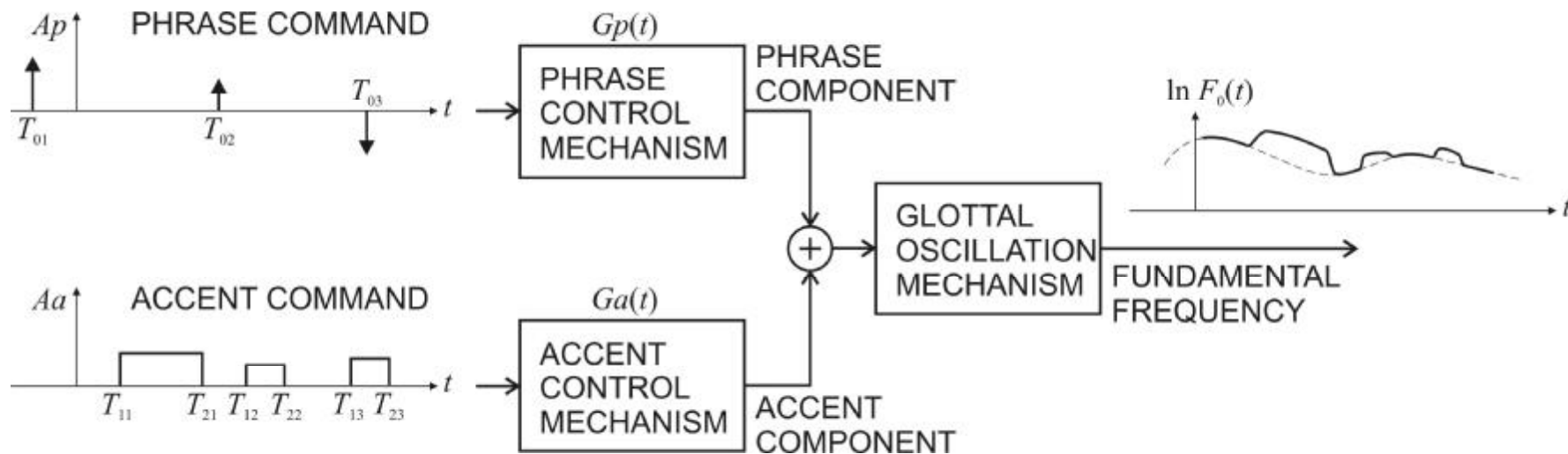
Modelování prozodie

- Pomocí pravidel
- Umělé Neuronové Sítě
- Prozodické modely

Intonační modely

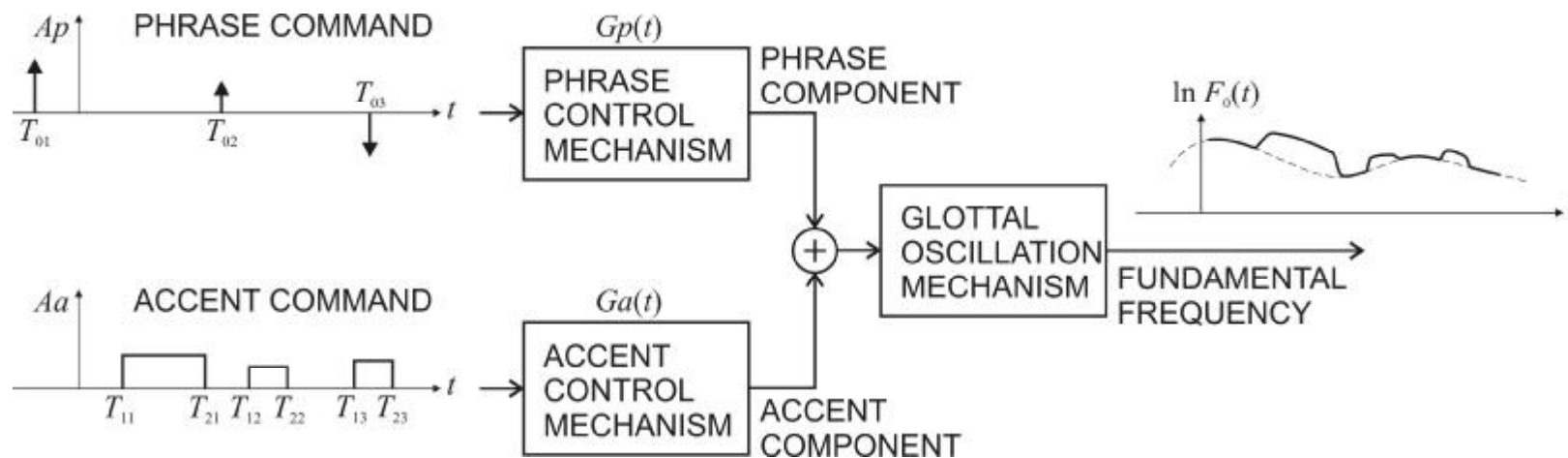
- Pierrehumbertová (1983)
- ToBI (1992)
- IPO (1990)
- INTSINT (1991)
- Fujisaki (1984)

Fujisakiho intonační model



$$\ln F_0(t) = \ln Fb + \sum_{i=1}^I A p_i G_p(t - T_{0i}) + \sum_{j=1}^J A a_j \left[G_a(t - T_{1j}) - G_a(t - T_{2j}) \right],$$

Fujisakiho intonační model

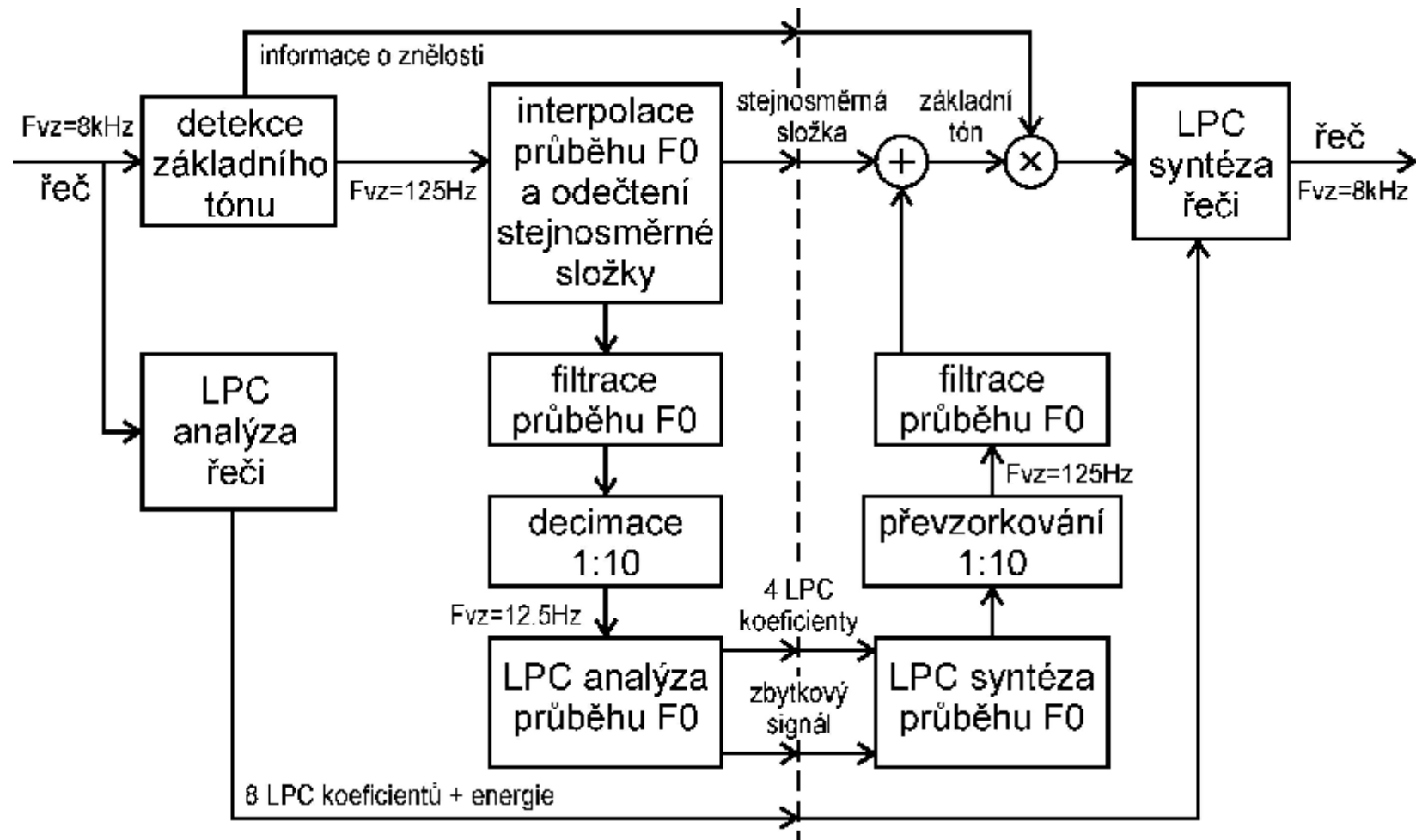


$$\ln F_0(t) = \ln Fb + \sum_{i=1}^I A p_i G_p(t - T_{0i}) + \sum_{j=1}^J A a_j \left[G_a(t - T_{1j}) - G_a(t - T_{2j}) \right],$$

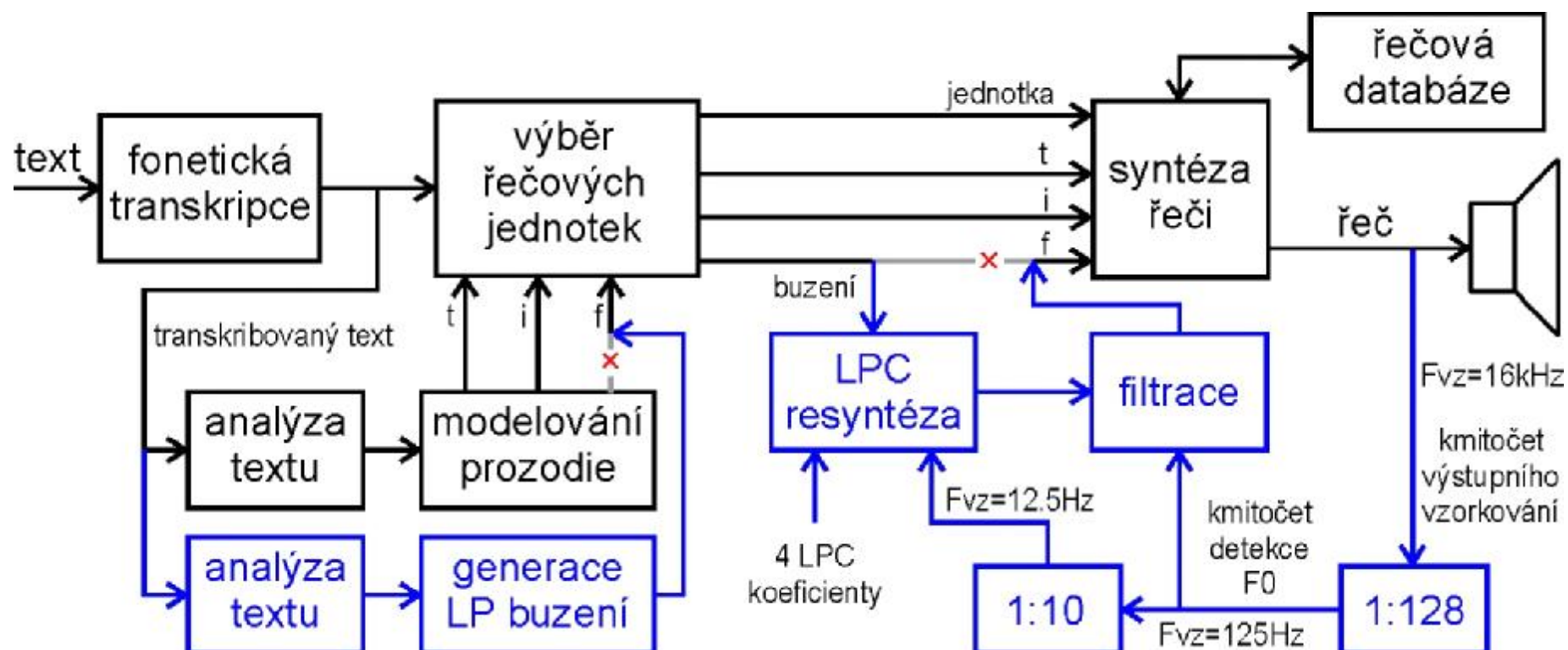
kde

$$G_p(t) = \begin{cases} a^2 t \exp(-at) & \mathbf{K} \text{ pro } t \geq 0, \\ 0 & \mathbf{K} \text{ pro } t < 0, \end{cases} \quad G_a(t) = \begin{cases} \min[1 - (1 + bt) \exp(-bt), g] & \mathbf{K} \text{ pro } t \geq 0, \\ 0 & \mathbf{K} \text{ pro } t < 0. \end{cases}$$

LP intonační model



TTS systém s LP intonačním modelem

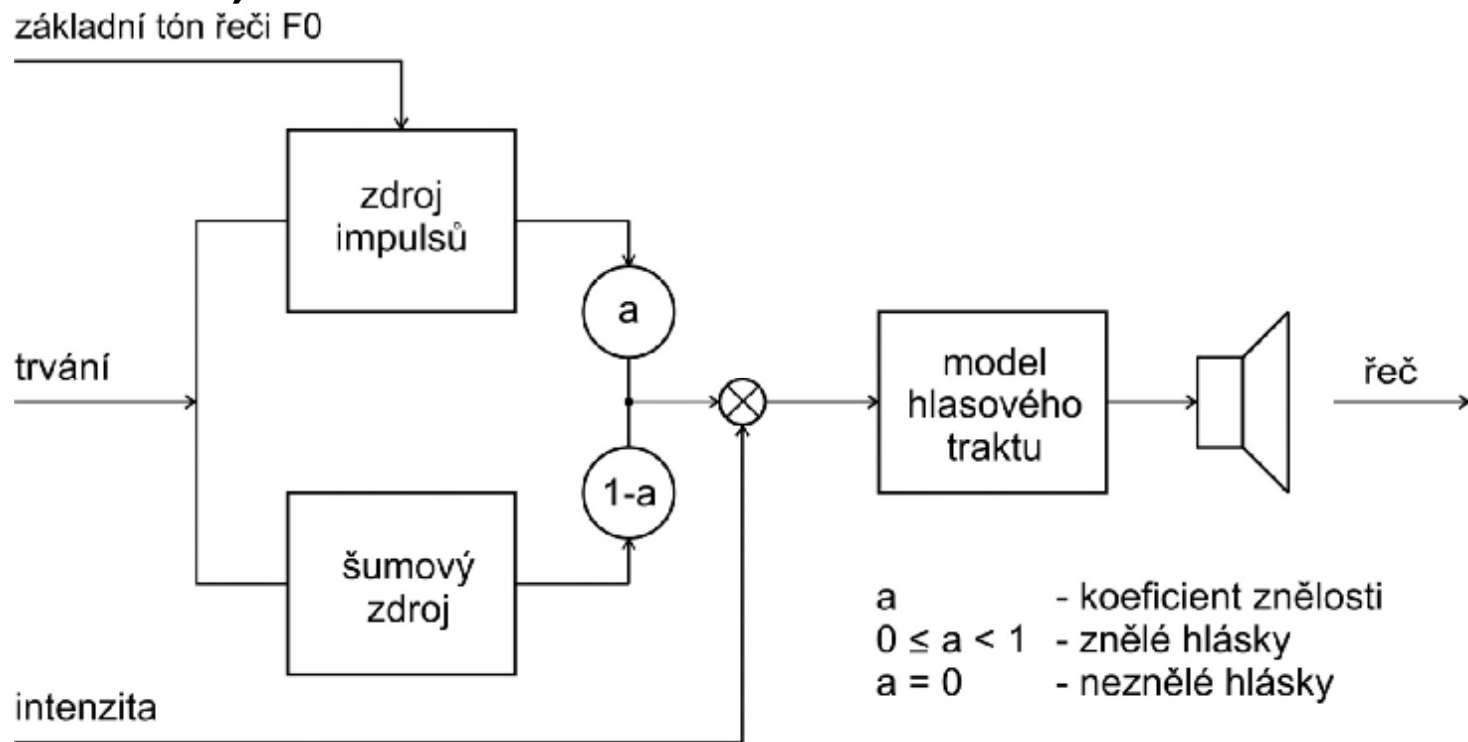


Metody syntézy řeči

- **Artikulační syntéza**
 - Modelování kompletního hlasového traktu
 - Výpočetně náročné, špatná kvalita
 - Prakticky se zatím nevyužívá, pouze ve výuce
- **Formantová syntéza**
 - Zjednodušené modelování hlasového traktu pomocí formantů a antiformantů
 - Praktické aplikace – nejpoužívanější v 60.- 80. letech
- **Konkatenační syntéza**
 - Řetězení segmentů řeči z inventáře řečových jednotek
 - Používá většina současných TTS

Akustická teorie vytváření řeči

- **Spojení dvou navzájem nezávislých složek**
 - Zdroj buzení (kvaziperiodický sled hlasivkových pulzů pro znělé zvuky, šum pro neznělé zvuky)
 - Lineární filtr (reprezentuje kmitočtovou odezvu hlasového traktu)



Formantová syntéza

- **Založena na akustické teorii vytváření řeči**
- **Zjednodušená simulace vytváření řeči člověkem**
 - **Zdroj buzení** – generátor impulsů pro znělé zvuky a šum nebo smíšené buzení pro neznělé zvuky
 - **Hlasový trakt** – modelování pomocí filtru, jehož parametry odpovídají formantům a antiformantům hlasového traktu
- **Syntéza podle pravidel** – parametry se nastavují podle manuálně nalezených pravidel
- **Dříve velmi úspěšná a používaná metoda**
- **Dnes se téměř nepoužívá (výjimka: DECTalk)**

Formantová syntéza

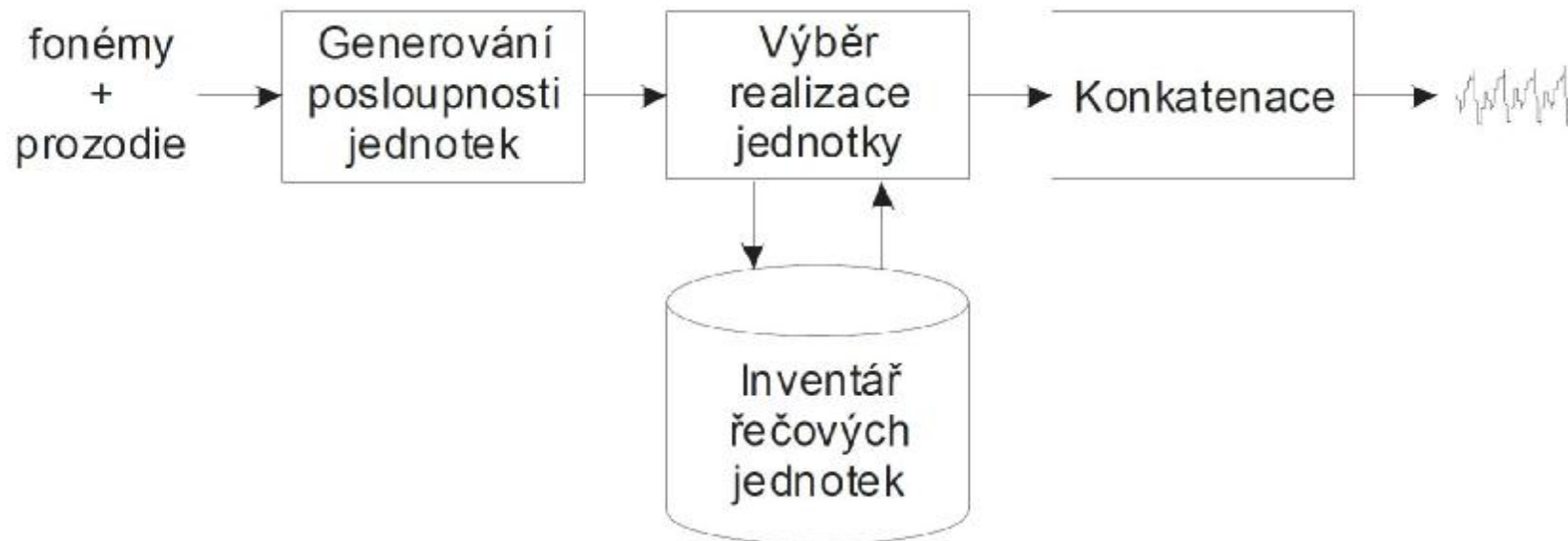
- **Výhody:**
 - Jednoduchý model
 - Snadné řízení prozodických charakteristik
 - Konstantní kvalita
 - Změny hlasu a emoce možno řídit podle pravidel
 - Schopnost vytvářet plynulou kvalitní řeč
- **Nevýhody:**
 - Pracné hledání a nastavování pravidel
 - Vzájemná interakce mezi hodnotami parametrů
 - Složitě vytváření některých zvuků (explozívy)
 - Nízká přirozenost řeči

Konkatenáční syntéza

- Používá přímo části přirozeného řečového signálu
- Předpokládá, že se řeč skládá z řečových (akustických) jednotek
- Řeč je možné rozdělit na segmenty odpovídající těmto jednotkám a uložit je do inventáře řečových jednotek
- Řeč se vytváří řetězením (konkatenací) řečových segmentů uložených v inventáři řečových jednotek
- Syntetická řeč napodobuje řečníka z inventáře

Konkatenáční syntéza

- Generování posloupnosti řečových jednotek
- Výběr vhodné realizace řečové jednotky
- Vlastní řetězení (konkatenace)
- Syntéza řízená daty – parametry syntetizéru se nastavují automaticky z řečových dat



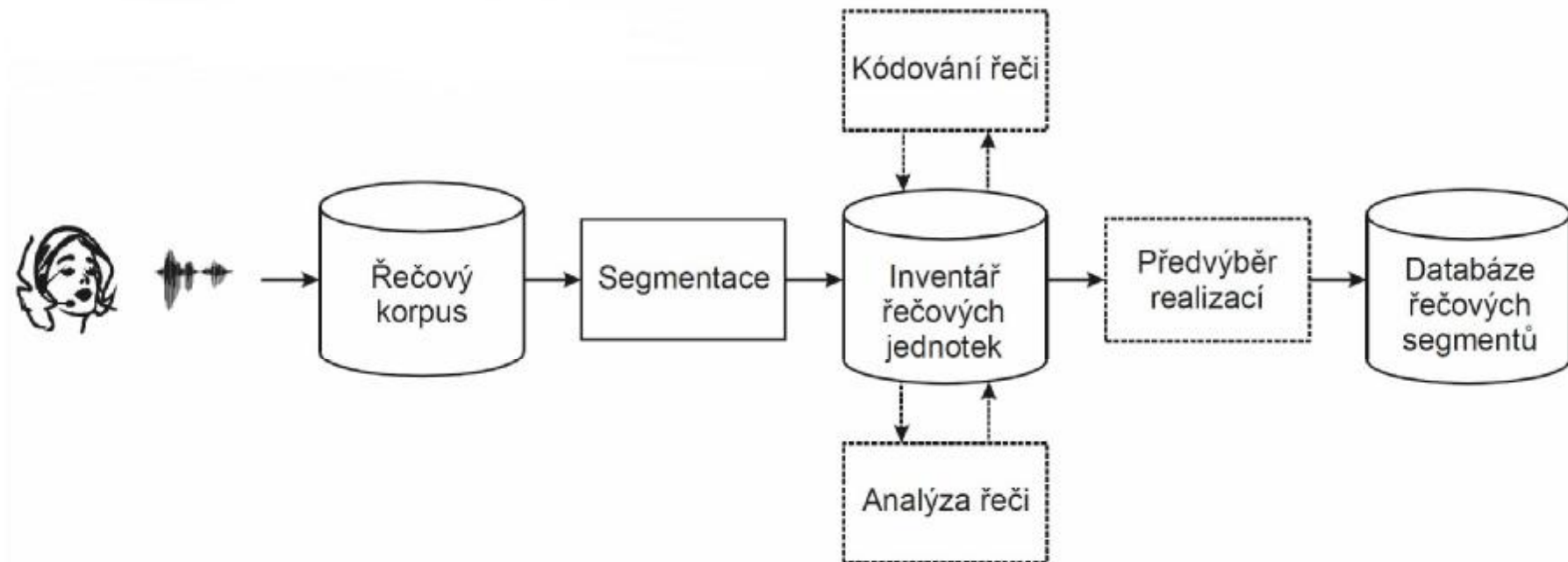
Konkatenáční syntéza

- **Vytváření inventáře řečových jednotek**
 - Ruční vytváření
 - Automatické vytváření
- **Způsob reprezentace řečových jednotek**
 - Neparametrická (přímo vzorky řeči)
 - Parametrická (LPC, keprální, HNM)
- **spektrální/prozodické modifikace jednotek**
 - Bez modifikací (pouze řetězení)
 - S modifikacemi (snaha o minimalizaci nespojitostí na hranicích řetězených jednotek)
- **Možnosti generování řeči**
 - S omezeným slovníkem (věty ze specifické oblasti)
 - S neomezeným slovníkem (libovolné věty)

Vytvoření databáze řečových jednotek

- Volba typu řečových jednotek
- Vytváření řečového korpusu
- Segmentace řečového korpusu
- „Předvybrání“ zástupců řečových jednotek
- Parametrizace řečových jednotek
- Kódování řečových jednotek

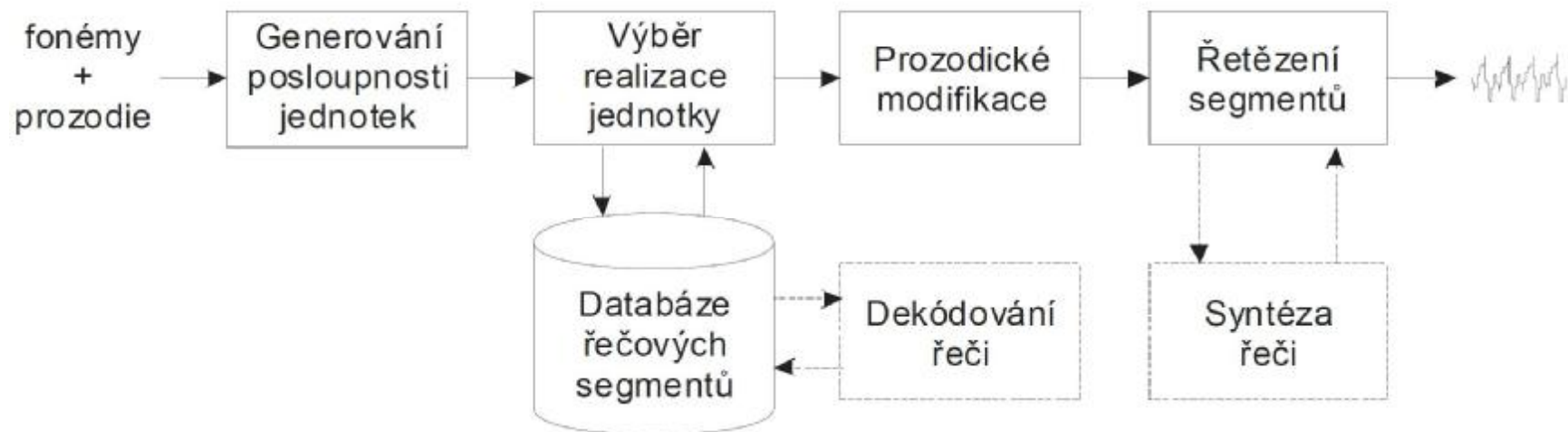
Vytvoření databáze řečových jednotek



Konkatenáční syntéza

- Posloupnost fonémů + prozodie
- Odvození posloupnosti řečových jednotek
- Výběr zástupce řečové jednotky z databáze
- Dekódování řečové jednotky
- Prozodické modifikace řečových jednotek
- Spektrální vyhlazování řečových jednotek
- Vytváření řeči na signálové úrovni
- Vlastní konkatenace

Konkatenáční syntéza



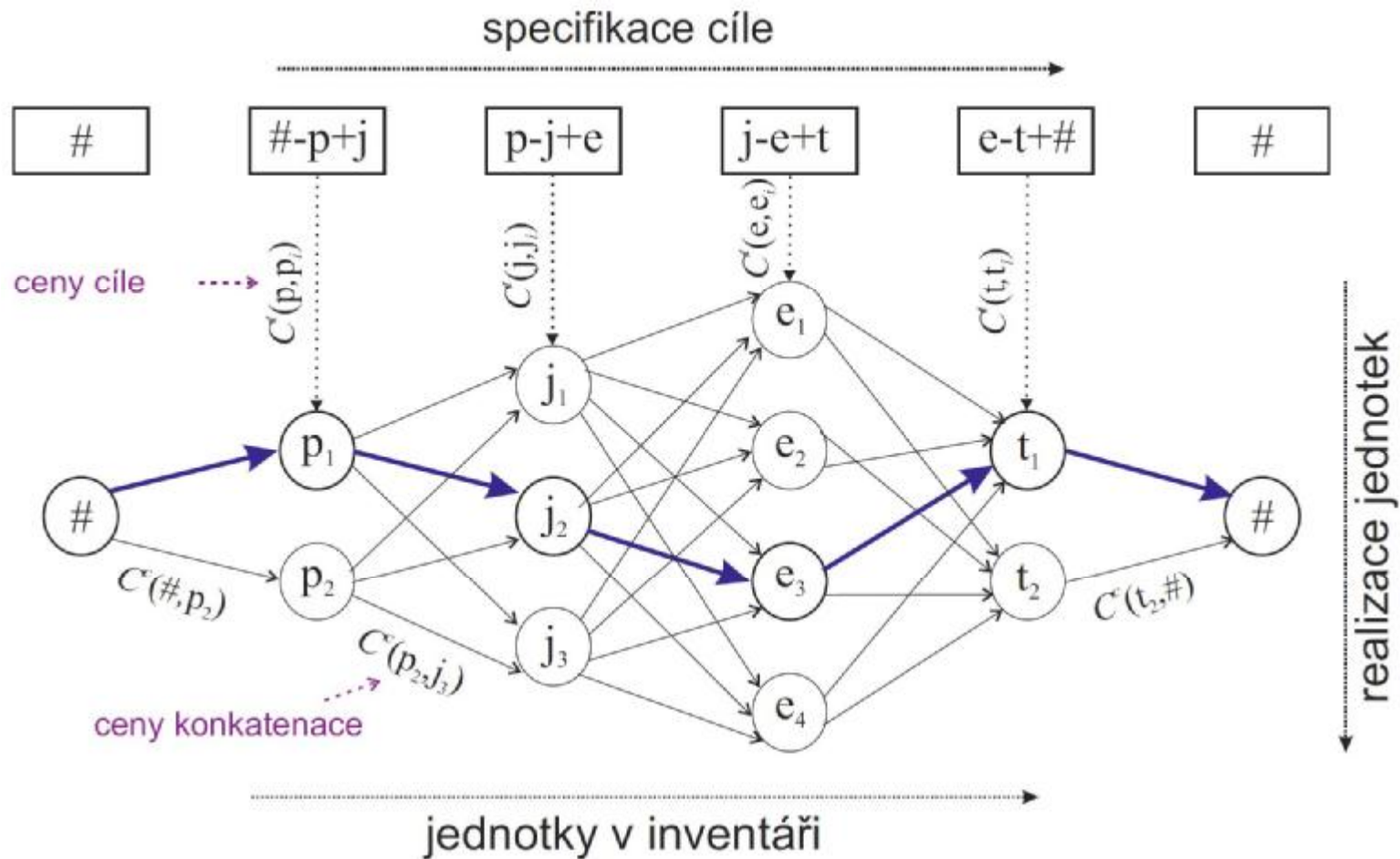
Korpusově orientovaná syntéza

- Zvláštní případ konkatenáční syntézy
- Využití rozsáhlých a prozodicky pečlivě anotovaných řečových korpusů
- Více realizací každé řečové jednotky (v rozdílných fonetických, spektrálních i prozodických kontextech)
- Plně automatická konkatenáční syntéza
- Všechny parametry se určují automaticky na základě dat z řečového korpusu
- Často tzv. neuniformní řečové jednotky (jednotky různého typu) –během syntézy se vybere typ a realizace jednotky
- => syntéza výběrem jednotek – Unit Selection

Korpusově orientovaná syntéza

- Hledání optimálních posloupností řečových jednotek (resp. Jejich realizací) v řečovém korpusu v rámci syntetizované promluvy
- Čím přesnější najdeme posloupnost jednotek, tím menší modifikace původního řečového signálu budeme muset provést => výsledkem je vyšší kvalita syntetické řeči
- **Dvě hodnotící funkce**
 - **Cena cíle C^t**
 - **Cena konkaténace C^c**

Korpusově orientovaná syntéza



Prozodické a spektrální modifikace

- Přiblížení prozodických a spektrálních vlastností vybraných zástupců řeč. jednotek vlastnostem požadovaných v syntetické řeči
- **Prozodické modifikace**
 - Úprava spektrálních vlastností řeč. jednotek z inventáře – přiblížení k požadovaným vlastnostem syntetické řeči
 - Plně v režii konkrétní metody
- **Spektrální modifikace**
 - Úprava spektrálních vlastností syntetické řeči
 - Dostačující většinou prostá lineární interpolace spektrálních parametrů (LPC, HNM)
- **Žádné modifikace** – teoreticky nejvyšší kvalita – žádná degradace řeč. signálu – potřeba gigantických inventářů
- **S modifikacemi** – větší pružnost systému – možno použít menší inventáře

Prozodické a spektrální modifikace

- **Metody**
 - Přímá syntéza
 - LP syntéza
 - PSOLA
 - Keprální syntéza
 - Harmonický a šumový model vytváření řeči (HNM)

Hodnocení kvality syntetické řeči

- **Kvalita**
 - Srozumitelnost, přirozenost, plynulost, příjemnost, přijatelnost uživatelem
- **Vzhledem ke komplexnosti řeči neexistují objektivní testy**
- **Poslechové testy – subjektivní hodnocení kvality**
- **Testy srozumitelnosti**
- **Testy přirozenosti**

Aplikace syntézy řeči

- Pomůcky pro nevidomé a slabozraké
- Telekomunikační služby
- Automatické čtení (SMS, e-maily ...)
- Hlasové monitorování
- Výuka jazyků
- Multimédia, komunikace člověk-počítač
- Mluvicí hračky pro děti
- Výzkum (fonetika, lingvistika, akustika)

České řečové syntezátory

- **Komerční**
 - ARTIC – ZČU Plzeň <http://www.speechtech.cz/>
 - Acapela Group – Eliška
 - RealSpeak - Zuzana
- **OpenSource**
 - Epos <http://epos.ufe.cz/> (zde jsou i zajímavé odkazy)
 - Festival (s podporou češtiny od Brailcomu)

Literatura

- J. Psutka, L. Müller, J. Matoušek, V. Radová: Mluvíme s počítačem česky. Academia 2006
- Ukázky k přednášce: <http://epos.ufe.cz/ukazky>